Using Social Media to Characterise Crowds in City Events for Crowd Management

Gong, X.

DOI
10.4233/uuid:e13606bc-e466-44c6-8a98-dd89aac8fdc4

Publication date
2020

Document Version
Final published version

Citation (APA)
Gong, X. (2020). *Using Social Media to Characterise Crowds in City Events for Crowd Management*. [Dissertation (TU Delft), Delft University of Technology]. TRAIL Research School. https://doi.org/10.4233/uuid:e13606bc-e466-44c6-8a98-dd89aac8fdc4

Using Social Media to Characterise Crowds in City Events for Crowd Management

Xun GONG

The research leading to this dissertation has received funding from the European Research Council under the European Union Horizon 2020 Framework Programme for Research and Innovation. It is established by the Scientific Council of the ERC Grant Agreement no. 669792 (Allegro).

Using Social Media to Characterise Crowds in City Events for Crowd Management

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof. dr. ir. T.H.J.J. van den Hagen
chair of the Board for Doctorates
to be defended publicly on
Monday 21st, September 2020 at 15:00 o'clock

by

Xun GONG
Master of Science in Computer Science
Delft University of Technology
born in Jiajiang, Sichuan

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof. dr. ir. S. P. Hoogendoorn	Delft University of Technology, promotor
Dr. ir. W. Daamen	Technische Universiteit Delft, promotor
Prof. dr. ir. A. Bozzon	Technische Universiteit Delft, copromotor

Independent members:

Prof. dr. ir. G. J. P. M. Houben	Delft University of Technology
Prof. dr. M. Brambilla	Politecnico di Milano, Italy
Dr. M. Gonzalez	University of California, Berkeley, U.S.
Prof. dr. ir. M. Van Steen	University of Twente

TRAIL Thesis Series no. T2020/14, the Netherlands Research School TRAIL

TRAIL
P.O. BOX 5017
2600 GA Delft
The Netherlands
E-mail: info@rsTRAIL.nl

ISBN: 978-90-5584-270-4

Printed in the Netherlands

Education changes lives.

Acknowledgement

When I was a master student in Computer Science, I often visited the Transportation and Planning department to play table tennis with friends. I didn't expect that I will continue my study here one day. However, only two months after my master defence, on Jun 20th 2016, I joined the Allegro project as a PhD student in this department. When I look back on my four years PhD journey, I felt discouraged and frustrated in the beginning, and was afraid of making mistakes. But now I feel relieved and have a huge sense of achievement. I believe this would have never been possible without the support and guidance from various people in the institution, colleagues, friends, and family.

First and foremost I would like to express my sincere gratitude to my promoter, Serge Hoogendoorn, for funding my PhD and supporting my research. Meetings with Serge always resulted in creative ideas and valuable suggestions, which inspired and motivated me a lot and to keep me on the right track. Serge, special thanks to your effort and quick response in reviewing my thesis multiple times amid your extremely busy schedule. The time and energy you spent with me were far beyond what I would have expected. Then my deep appreciation goes to my supervisor and co-promoter, Winnie Daamen. She is a very nice and kind person, but in the meantime also very critical in the work. Guidance by Winnie is an extensive experience where I felt gloomy at first, e.g. dealing with a large number of comments in a single document, but I gradually felt released when significant improvements in my work are recognized as a result of processing from her feedback. Finally, I feel grateful when I see improvements in myself in the way of thinking. I think this is what good education should be like. A unique and valuable experience worth having. Winnie, it is my honour to be your student. Furthermore, I would like to thank my supervisor and co-promoter, Alessandro Bozzon. I have worked with Alessandro for more than 6 years now, starting with the Information Retrieval course during my masters, followed by my master thesis, and continuing with my PhD research. Alessandro, I'm very grateful for all the fruitful discussions, insightful comments and wonderful suggestions you gave in these years. I am very impressed by your talent and knowledge.

Secondly, I would like to thank the independent committee members for reviewing my thesis and participating in my defence. Thank you, Marta C. Gonzalez, Marco Brambilla, Maarten van Steen and Geert-Jan Houben. Special thanks to Geert-Jan, for his support as my master thesis promoter and at the beginning of my PhD research.

Thirdly, I would like to thank all ALLEGRO project members: Lara-Britt Zomer, Danique Ton, Florian Schneider, Alexandra Gavriilidou, Tim van Oijen, Ilse Galama, Giulia Reggiani, Alphonse Vial, Martijn Sparnaaij, Marie-Jette Wierbos, Yan Feng and Dorine Drives. Together with all of you, we explored interesting topics and created lots of fun. Lara, the spirit of our office, I appreciate all your creative ideas, which really helped a lot. Danique, what? I believe one day you will be a professor. In your word, ”Luctor et Emergo”. Florian, yes I'm back! So glad to have you around to have lots of fun, not only in the office, or at the coffee machine, but also on the football pitch. Many thanks for taking care of the DCF Inter when I'm not there. Alexandra, thanks for helping me out whenever I needed it. I promised to write a piece of code to close the windows in the office automatically once you enter. Tim, it was very pleasant working with you on those interesting programming tasks. Giulia, when I was at the school I was motivated by saying 'All Roads Lead to Rome'. But you were born in Rome. What should I say? Such an unfair world. Thanks for the funny moments you created. Alphonse, thanks for your open ideas, professional skills and help, not only as project mates but also during the football matches. Martijn, our professional goalie and defender, thanks for your help in the office, and your consistent performance in the matches. Marie-Jette, thank you very much for your help, for instance, translating Dutch phone calls for me several times. Yan, welcome to join this office. It's pleasant to work with you in those collaboration tasks. Dorine Drives, many thanks for your suggestions on my research at the beginning of my PhD.

Additionally, I would like to thank all colleagues inside and outside the department. It was nice meeting all of you and I enjoyed our talks during meetings, lunches and coffee. Thank you, Haneen Farah, Andreas Hegyi, Victor Knoop, Goncalo Correia, Rob van Nes, Niels van Oort, Egidio Quaglietta, Ehab Al Khannaq, Paul van Gent, Jeroen van der Gun, Xiao Lin, Kees van Goeverden, Louise Klingen, Michela Le Pira, Maria Salomons, Simon Goede, Edwin Scharp, Peter van Oossanen. Joelle Aoun, Na Chen, Malvika Dixit, Ximing Chang, Nejc Gerzinic, Jeroen van den Heuvel, Panchamy Krishnakumari, Freddy Mullakkal Babu, Hari Nagalur Subraveti, Jishnu, Narayan, Solmaz Ramzi Rad, Nagarjun Reddy, Sanmay Shelat, Yaqing Shu, Silvia Varotto, Fei Yan, Guopeng Li, Xiaocheng Ma, Bart van Arem, Sascha, Flurin Hanseler, Ilse Galama, Yu Han, Maryna Ozturker, Tin, Bahman, Pablo, Peyman, Nikola, Dehlaila, Priscilla, Nicole, and Moreen. Special thanks to Paul van Erp, I will never forget how you saved me by lending your shirt for my TRB presentation. Maybe I should count you as a co-author. Thanks, buddy. Yufei & Xiao, thanks all for the help and support in the past years, including drinks, talks, and BBQs. Maria, the trip to NY was so nice, particularly the show we watched. Thanks for arranging it, and also thanks for the tips and suggestions you have given about graduation. Meng Wang, I wonder how I can be fit as you are when you are playing football but are also doing wonderful researches? Magic. Niharika, nice to have you around. You bring so much fun. Xiao & Qu, Lin & Zheng, Yaqing & Lan, Pengling & Song, glad to meet you all during my PhD, and thanks for all your help and support. It was fun gathering with you all. Konstanze, thanks so much for the spicy sauce and the wonderful party. Yongqiu & Hongrui, the only ones with whom I can practice the same dialect as speaking to my parents but then at the university. Many thanks for the help in the past four years. Alessandro Bombelli, thanks for the discussion about clustering analysis. Jing Zhao, a nice lunch buddy, we should do this more often. Oded Cats, thanks for joining my go-not-go meeting. To the exchange PhD students group and friends, i.e. Summer, Wenjing, Rongqin, Yu, Zhuo, Shi, YY and ZY, I'm so lucky to meet you all. The wonderful road trip to France was one of the most precious moments during my PhD. Cheers! Special thanks to Meiqi, for all those good memories which I shall never forget. Special thanks to Ding and Yihong, besides all the fun we created, my deep appreciation for all those critical and meaningful discussions, as well as wonderful hangouts and travels.

In the meantime, I would like to thank colleagues who worked together in the Alessandro Bozzon's team, i.e. Jie, Sepideh & Shahin, Siyu, Carlo, Ziyu, Andrea and Achilleas. Special thanks to Jie, my co-author, also my supervisor in my master thesis. Jie, It was a pleasure to work with you alongside your extraordinary guidance and contribution to our work.

One of my favourite sports is football, where I met a large number of friends. I joined two teams, DCF and DCF Inter. Hereby I would like to thank all friends, teammates, superstars in these two teams. Without you, I would not have enjoyed my life here at the university. Special thanks to Di María (Jiakun), Fan, Yaxun, Xiaoqian, Kelu, Junda, Hao, Jinku, Andi, Simo, Zheng L, Li W, Zhaoxin, ZLY, Zhu, Jian, Qiang, Luke, Xiaocheng, Zequn, and Kai in the DCF. With regard to the DCF Inter, special thanks to Florian, not only for your wonderful goals but also for the support during team building. Alphonse, I still remember the first day when I met you and invited you to join our team. You are so professional! Martijn, Tim and Alessandro, legend goalie and defenders. Alessandro Bombelli, my Italy bro, thanks for your splendid support not only in clustering analysis but also the defences in the football matches. Nikola, running faster than the wind, always tackling the ball and scoring unexpected goals. Pablo, stable and effective defence, thank you so much for inviting your brother joining our games. Xavi, the legend from Barcelona. Thanks for travelling so long to join the games on Monday evenings! Bahman, you are so fast and always attract more than two defenders. Oguz, like a wall in front of opponents, thanks for your marvellous defending and attacking. Sako, our Japan star with rocket speed! Jeanette, Junda (Tiger), Wangzi and Chang, thanks for joining the games now and then.

Moreover, I would also like to thank all the friends I met and hangout during these years. With you, every moment counts. Thank you, Roland CS group, 2624AV roommates and S&S group, Yellow Spirits assignment group (Yen, Sara, HH). Special thanks to Yen for translating the summary to Dutch and reviewing my English. Special Thanks WLD for those wonderful movies, music and food. Special thanks to Ida, Xiaopeng, Xinyang, Kexin, Chen, Yuteng, Feng, Ge, Mhh, Xjx, Qi, Shumi, Zheng, Binlai, Junzhi, Jiayun, Naive, Yiwen, Dichao, Yunyao, He, Rong, Canran. My deep appreciation goes to GL, Xi and LC, for sharing ideas and having meaningful discussions. I'm eternally grateful for your understanding and support.

Besides, there are friends and colleagues I would also like to acknowledge in my previous study and work. Special thanks to Nathan, Meg, Litte east, Pan & Feng, Patrick, Linna, Fu, Gavin, Fan, Lei, Yajun & Liang, Yili, Zonghui, Jie Z, Veka, Haixia Chi, Jing Chen, Zheng Wen, Qi Chen, Eelco, Sue Wu, Xiaobin, Zhongyu, Wei, Hua Tang, Lei L, Axin, Bo, Liu, WW, QW, Tinghong, Lin P., Jing S & Xiang, Jin L, Jie Y, Jia C, Daohong, Qian, SJ, Geng, Kai, Shuoyan & Juan, SY, Lxt, Jun, Rong, Fish, Yi, En, Maosheng, Yi & Jia, Xuejuan & Hui, Lan. Special thanks to CD group, i.e. Mei, SX &

Xiaoqin, Lincat, Shuqian & Yang, ZQ, Rui, Sha, P4, Jun, Hui, and Yuan.

Most important, I would like to express my eternal appreciation to relatives and family members. Thank you, Enguo, Yuhua, Yi, Liang, Yan, Minghua, Yihui, Ying, Eva, Tongtong, Delong, Mingshu, Min, Feng, Zhi, Lihui, Yihuang, Liu, Xiaoya, Zuokun, Liangcheng. My utmost appreciation to my parents for supporting me pursuing my dream while breaking the shackles of tradition. I would have never made it without your constant support.

Finally, this thesis is dedicated to the people in the past and in the future who devote themselves to the struggle for democracy, freedom and equality, for a better life.

X. Gong

Delft, April 2020

Contents

Preface **vii**

1 Introduction **1**

1.1 Research scope 3
1.2 Challenges in using social data 3
1.3 Research objective and research questions 4
1.4 Main contributions 5
1.4.1 Scientific contributions 5
1.4.2 Practical contributions 7
1.5 General research approach 7
1.6 Outline 9

2 Crowd Characterization for Crowd Management using Social Media Data in City Events **11**

2.1 Introduction 13
2.2 Crowd Characterization 15
2.2.1 Characterization factors 16
2.2.2 Social Media Data Analysis for Crowd Characterization 18
2.3 Applying crowd characterization based on Social Media data in two city-scale events 22
2.3.1 Case selection 23
2.3.2 Data collection 26
2.4 Findings & Analysis 28
2.4.1 Demographic analysis 28
2.5 Summary and conclusions 43
2.6 Appendix 45

3 Estimate Sentiment of Crowds from Social Media during City Events 51
3.1 Introduction . . . 53
3.2 Literature Review . . . 55
3.3 Research Approach . . . 57
3.3.1 Comparison metrics . . . 58
3.4 Selection of Candidate Methods from Literature . . . 59
3.4.1 Lexicon based methods . . . 59
3.4.2 Machine Learning based methods . . . 61
3.4.3 Sentiment estimation result scheme . . . 62
3.5 Data Collection . . . 63
3.5.1 Common-based dataset . . . 64
3.5.2 Event-based dataset . . . 65
3.6 Experimental Setting . . . 69
3.6.1 Control variable . . . 69
3.6.2 Scenario design . . . 70
3.6.3 Experimental setting . . . 70
3.7 Sentiment Analysis: Findings of the Experiment . . . 73
3.8 Discussion . . . 74
3.9 Conclusion . . . 75

4 Counting People in the Crowd Using Social Media Images for Crowd Management in City Events 77
4.1 Introduction . . . 79
4.2 Definition of Crowd size . . . 82
4.2.1 Crowd size levels . . . 82
4.2.2 Number of people in the crowd in less populated images . . . 83
4.3 Research approach . . . 85
4.3.1 Comparison metrics . . . 86
4.4 Social media image characteristics . . . 87
4.4.1 Crowd management perspective characteristics . . . 89
4.4.2 Social media perspective characteristics . . . 91
4.5 Selection of crowd size estimation methods . . . 92
4.6 Data collection and annotation . . . 95
4.6.1 Event selection . . . 95
4.6.2 Social media data collection . . . 95
4.6.3 Data annotation . . . 96

4.7 Experimental setup 101
4.7.1 Experiment 1: Crowd size estimation accuracy . . . 101
4.7.2 Experiment 2: Impact of image characteristics on crowd size estimation 101
4.8 Crowd counting analysis: findings of the experiment on crowd size estimation and image characteristics impacts 102
4.8.1 Crowd size estimation from social media images in city events 103
4.8.2 Impact of image characteristics on crowd level estimation 107
4.9 Discussion 110
4.10 Summary and Conclusion 111

5 Using Social Media for Attendees Density Estimation in City-Scale Events 113
5.1 Introduction 115
5.2 Related Work 117
5.3 Estimating Attendees Density from Social Media Data . . . 118
5.3.1 Pedestrian Traffic State Variables 119
5.3.2 Geo-based Density Estimation 120
5.3.3 Speed-based Density Estimation 121
5.3.4 Flow-based Density Estimation 123
5.4 Experimental Setup 124
5.4.1 Event selection 124
5.4.2 Experimental Conditions 130
5.4.3 Data Collection 130
5.5 Results 134
5.5.1 Results of Density Estimation 135
5.5.2 Δt Sensitivity Analysis 140
5.6 Discussion 141
5.7 Conclusions 145

6 Conclusions, implications and recommendations 149
6.1 Main findings 149
6.2 Overall conclusions 153
6.3 Implications for crowd management practice 155
6.3.1 Adopt Web-based data sources in the crowd management operations 155

6.3.2 Insight into biases introduced by social media sources and/or content analysis techniques 155
6.3.3 Promote social media usage in city events for crowd management 156
6.4 Recommendations for future research 156
6.4.1 Address geo-referenced social media data sparsity . 157
6.4.2 Novel data-fusion techniques 157
6.4.3 Gaining experience by applications during running events 158

Bibliography **159**

Summary **179**

Samenvatting (Summary in Dutch) **185**

Summary in Chinese **191**

About the author **195**

TRAIL Thesis Series publications **197**

Chapter 1

Introduction

City-scale events are getting more popular and attract a large number of people participating in various activities. For instance, on King's Day, a national holiday in the Netherlands, a huge amount of people pour into the city and gather in the urban area, participating in various activities such as street parties, music festivals and boat parades. Event stakeholders, such as event organisers, police, municipalities, and crowd managers manage the crowd to avoid incidents. Crowd management practice consists of two phases (Martella et al., 2017), i.e. the planning phase and operational phase. In the planning phase, crowd managers require the past event data to infer guidelines and to perform computer simulations of the crowds in the event. In the operational phase, a set of what-if scenarios are resembled. Crowd managers require the current information of the crowd to decide which scenario is met and to predict and prevent accidents. Further, they can apply feasible predefined measures in the given scenario. The information required either in the planning phase or operational phase includes the information of the crowd about a set of factors that influence pedestrian behavior (Martin, 2006; Zomer et al., 2015) and crowd management (Li, 2019; Still, 2000; Tubbs & Meacham, 2007; Abbott & Geddie, 2000). Examples of these factors are crowd size, density, mobility, emotion, visitor profile and location, which will be further discussed in Chapter 2.2. Conventionally, the information about these factors is derived from data collected by existing resources, such as stewards and ICT solutions based on sensors, e.g. Bluetooth/ Wi-Fi sensors, GPS trackers and Video cameras.

Existing resources used to acquire crowd information have drawbacks. Table 1.1 lists a set of factors which information about these factors are required by crowd management and show the overview of such drawbacks.

According to Table 1.1, stewards can report information about all required factors, but they are expensive and prone to bias. ICT solutions based on sensors collect Spatio-temporal data about the crowd, which can be used for deriving information about influencing factors. However, they are not broadly adopted, and can not provide semantic information (e.g. emotion). Besides, accessing the images or video recordings of public areas is computationally intensive, and often restricted due to privacy issues. These disadvantages reduce the capability of existing solutions providing information about influencing factors for crowd management.

Table 1.1: The capability of existing solutions and social media in providing information about factors influencing event planning for crowd management.

Factors	Inquiries	Required Information	Existing solutions				Social media
			Stewards	Bluetooth/ Wi-Fi	GPS	Video	
Visitor profile	What is the demographic of the people in the crowd?	Demographics, city-role	L	-	-	L	G
Crowd size	What is the number of people in the crowd?	Number of people	L	L	L	L	G
Density	Is the density of the crowd high?	Density	L	L	L	L	G
Mobility	What is the position of the crowd in different moments?	Temporal˜distribution, position	L	L	L	L	G
Location	Which PoIs are popular during the event?	Position, Points of Interest	L	L	L	-	G
Semantic	Is the crowd in intensively negative sentiment?What are the topics the crowd talking about?	Sentiment,˜word use	L	-	-	-	G

L = Locally, G = Globally, - = Impossible.
Factors are screen out according to pedestrian behavior (Martin, 2006; Zomer et al., 2015) and crowd management (Li, 2019; Still, 2000; Tubbs & Meacham, 2007; Abbott & Geddie, 2000)

Next to these traditional data sources, the advent of web-enabled technologies provided new sources of social data that could be used to analyse and understand human behaviour at large events. On social media, a user sends a post with content and profile at a certain timestamp. A certain percentage of posts is also tagged with geo-referenced information, i.e. the location where the post is sent. Compared with conventional solutions, i.e. stewards and ICT solutions, social media data is created by the people themselves, enriched with Spatio-temporal annotations, and integrated with rich semantic descriptions. Besides, social media data is created at high frequency, and it is free to access. At the same time, there are some disadvantages in social media data analysis. First, it is sparse, in particular for the geo-referenced social media posts. It is also biased, in the sense that there is a mismatch between the composition of crowds and the composition of social media users. Regardless of the disadvantages, social media is a promising source of knowledge for deriving characteristic information about the crowd.

We therefore argue that social media data can enrich considerably the information needed for crowd management, both in the operations phase and in the planning phase. In this thesis, we investigate to what extent social media are able to provide information for crowd management in city events in terms of three topics, i.e. event characterisation, attendees' sentiment and density estimation.

1.1 Research scope

There are various types of social media platforms. In this research, we focus on social media platforms on which social media data are globally available and can be accessed freely through an API or through a web page. The data of social media should contain either or all of the following attributes: the profile of the user who sent the post, the content of the post which can be text and images (optional), the timestamp when the post is sent, and the geo-location where the post is sent (optional). We collected social media data from different platforms on which the social media data are in different structures. We then converted them into a unified structure (Ou, 2011) which contained these four elements. We perform our research in the context of city events, which take place in cities, with a start-time and end-time and lasts for one day to several days.

1.2 Challenges in using social data

Though social media data are publicly available, the efficient and effective use of social media data to derive information about the crowd for crowd management is still unclear. There are three major challenges to be addressed in this thesis.

The first challenge is how to deal with the sparsity of social media data, both in time and in space. According to previous research (Paule et al., 2019; Middleton et al., 2018), the geo-referenced social media posts collected from social media platforms only account for a small proportion of the total messages posted. Particularly after the #noGeo activities on Twitter and Instagram (Holson, 2018; Dickinson, 2019), the availability of geo-referenced social media data is further reduced. This is further discussed in Chapter 5.1.

The second challenge is the bias of social media data. Recent works (Ribeiro et al., 2018; Yang et al., 2016; Duggan & Brenner, 2013; Díaz et al., 2018) show that social media is not used equally. Bias exists in social media usage in terms of, for instance, different age groups (Díaz et al., 2018; Yang et al., 2016), i.e. younger generation use social media more than older ones; different gender (Ribeiro et al., 2018; Duggan & Brenner, 2013; Yang et al., 2016), i.e. female are more fond of image-based social media such as Instagram than male. Similar to the sparsity of social media data, the bias of social media data reduces the capability of providing an accurate reflection of reality in city events. Thus, it increases the difficulties in deriving crowd information for crowd management.

The third challenge is how to derive information about the crowd using social media data (Feng et al., 2019; Ghani et al., 2019; Bocconi et al., 2015; Titos Bolivar, 2014) for crowd management. The information about crowds does not exist directly in social media data. Instead, social media data is full of noise, which may reduce the accuracy of information derived from social media. How to establish methods under such conditions is essential for deriving this information about the crowd for crowd management.

According to the major challenges introduced above, to derive information about crowds in city events that is in line with the reality for crowd management, we have to overcome these challenges.

1.3 Research objective and research questions

According to factors listed in Table 1.1, we screen out a list of information about these factors which are required by crowd management and possible to be derived from social media (Section 2.2.1). They are: demographic composition, city-role composition, Spatio-temporal distribution, Points of Interest preferences, word use, sentiment estimation, crowd size and density estimation. The main objective of this research is to understand how social media can be used as the data source to derive these information about the crowd for crowd management in the context of city events.

To achieve this objective, four research questions will be answered in this thesis:

RQ1. To what extent are social media data able to characterize crowds in city events, in terms of demographic composition, city-role composition, Spatio-temporal distribution, Points of Interest preferences and word use? (Chapter 2 of this thesis)

RQ2. To what extent are social media data able to estimate the sentiment of crowds in city events? (Chapter 3 of this thesis)

RQ3. To what extent are social media images able to count people in city events? (Chapter 4 of this thesis)

RQ4. To what extent social are media data able to estimate the density of people in city events? (Chapter 5 of this thesis)

1.4 Main contributions

The section below describes the contributions of this research, which are categorised into scientific contributions and practical contributions.

1.4.1 Scientific contributions

The scientific contribution consists of three types of contributions, i.e. methodological contributions, novel insights and constructed datasets.

Firstly, the methodological scientific contribution is the new density estimation method (Chapter 5), based on pedestrian traffic flow theory adapted to social media data analysis. The new model consists of three classes of density estimation strategies, namely: 1) geo-based strategies, operating only on social media data; 2) speed-based strategies, which estimate density by considering the travel speed (i.e. distance covered per unit of time) of attendees on the event terrain; and 3) flow-based strategies, which consider travel flow information (i.e. the number of attendees passing a reference point per unit of time). This research shows the successful application of pedestrian traffic flow theory in this new density estimation method.

Secondly, the social media data analyses are performed in this thesis as case studies (Chapter 2 to 5). The data analyses provide a better understanding of how social media data can be used to derive information about crowds, in terms of various factors listed in Table 1.1, required by crowd management in city events. The data analysis helps crowd managers to apply feasible and effective predefined measures to manage the crowd in different situations.

Lastly, a set of social media datasets are collected and annotated in this research (Chapter 2 to 5). The datasets are collected from various city events, which are selected considering diversities in various aspects, such as different cities, areas, event characteristics, editions, and activities. The collected datasets include Sail (2015), King's Day (2016, 2017, 2018), and Europride (2016, 2017) in the city of Amsterdam, and Feyenoord football fan riots (2017) in the city of Rotterdam. The annotated datasets consist of two parts, namely: 1) the social media posts annotated with the number of people in the images and image characteristics; 2) the social media posts annotated with sentiment based on the text. These collected and annotated datasets can be further used for, for instance, developing and verifying new models, studying and analysing cases.

The contributions described above are listed in the following.

CB1. Methodological contribution

- The density of participants estimation method

CB2. Novel insights

- City events characterisation in Sail 2015 and King's Day 2016 Amsterdam, in terms of:
    * demographic composition
    * city-role composition
    * spatio-temporal distribution
    * crowd size estimation
    * Points of Interest preferences
    * word use
- Density of participants analysis, in Sail 2015 and King's Day 2016, Amsterdam

CB3. Constructed datasets, collected from Twitter and Instagram, annotated with sentiment and crowd size

- SAIL (2015)
- King's Day (2016, 2017, 2018)

- Europride (2016, 2017)
- Football riots (07-05-2017)

1.4.2 Practical contributions

The findings reported in this work also have important practical contributions.

The developed density estimation model, together with verified existing algorithms, serve as a new method using social media data to derive information about people in the crowd for crowd management. For crowd managers, it is possible to select suitable measures to manage the crowd based on the derived information.

As mentioned before, the data analyses provide insights in crowds in city events, in terms of demographic composition, city-role composition, Spatio-temporal distribution, crowd size estimation, Points of Interest preferences, sentiment estimation and word use. In addition to helping crowd managers select feasible measures to manage the crowd, these insights also lead to recommendations to the event organisers, city administrators and urban planners, such as on improvements of bottlenecks in the event area and traffic hub.

Besides, the constructed pipelines as tools can be used by practice for future research, such as crawl data from various social networks, process data, analyse and visualise data.

1.5 General research approach

In this research, we employed a set of steps as a methodology framework to study four research questions. In each study, this research framework is applied differently. In the following, we first introduce the general steps in the methodology framework and further describe how these steps are performed in each study.

The methodology framework consists of four steps, as shown in Figure 1.1, i.e. literature review, methods comparison, case study and analysis. In the first step, we review the literature about deriving crowd characteristics information using social media data. Further, we compare existing state-of-the-art methods from literature. In case there are no feasible methods for tackling challenges, we propose new methods. To assess the effectiveness of

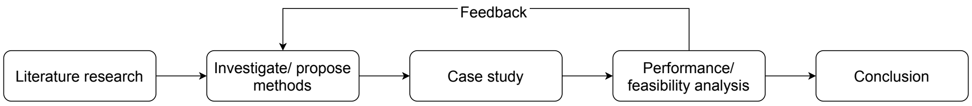


Figure 1.1: Research methodology framework that has been applied in investigating for each of the research questions.

existing state-of-the-art methods and proposed new methods, in the third step we conduct case studies, applying the methods to derive crowd characteristics from social media dataset collected during cases (city events). Next, we analyse the performance of methods by comparing the derived information with ground truth estimated from various data sources.

For the first research question, we investigate to what extent can social media be used to characterize crowd in city events in terms of various aspects. To decide which aspects of information are required for crowd management, based on Table 1.1, we perform literature review and screen out a set of factors that are related to crowd management. To retrieve crowd information about these factors, we review recent works to screen out a set of state-of-the-art methods. We investigate the effectiveness of these methods in a case study, using collected social media data in two city events to assess these state-of-the-art methods compared with estimated ground truth.

For the second research question, we study how social media can be used for estimating the sentiment of crowds in city events. To estimate the sentiment of social media texts, we screen out various types of state-of-the-art methods from recent works, and apply social media datasets collected in diverse city events on these methods. The ground truth sentiment of these social media texts are annotated through crowdsourcing. The effectiveness of these methods is calculated by comparing the estimated sentiment with the ground truth.

The third research question, i.e. investigating the methods for counting people in the crowd, is studied in a similar way to the second research question. We select state-of-the-art methods and perform case studies with annotated datasets. Instead of annotating sentiment of the post texts, in this study we construct a dataset by annotating the crowd size in the social media images, which consists of 1) the dense level of people in the image, and 2) the specific number of people in the image if less than 20. For each image, we also annotate the value of a set of image characteristics, e.g. if a picture shows indoors or outdoors, whether it is a selfie or a group picture. In the analysis step, in addition to analysing the effectiveness of the selected methods, we also analyse the impact of these image characteristics on the effectiveness of different methods in counting people in the crowd from social media data.

In the last research question, we study the density estimation of crowds in city events using social media. According to the literature review, there are no existing methods feasible to this research. Thus, we propose a new method to estimate the density of people in the crowd. The proposed method is based on pedestrian traffic flow theory and adapted to be used with social media data. It considers three strategies, i.e. geo-based, speed-based, and flow-based, using social media data. To analyse the performance of different strategies in the proposed methods, a case study is performed in five terrains (area) in two city events. The results from the proposed method are further compared with the ground truth estimated from sensor data.

1.6 Outline

The remainder of this thesis is organised in 6 chapters. The schematic overview of these chapters with their relationships is shown in Figure 1.2.

Chapter 2 showcases the state-of-the-art methods using social media in city events to derive information about the crowd, in terms of demographics, city-role, Spatio-temporal distribution, Points of Interest preference, and word use. Research question RQ1 is answered in this chapter.

To further derive emotion of the crowd using social media rather than only word use, in Chapter 3 we perform a sentiment analysis of the people in the crowd based on text collected from city events. In this chapter, we construct a sentiment annotated dataset and validate a set of state-of-the-art methods using the constructed dataset. The result answers the second research question (RQ2).

Chapter 4 and 5 investigate estimating the size and the density of people in the crowd using social media in city events. Chapter 4 focuses on assessing the effectiveness of existing methods in counting the number of people using social media images. We also investigate the impact of a set of image characteristics on the people counting performance. The result of this chapter answers the research question RQ3. In Chapter 5, we propose a new method to estimate the density of people in the crowd using social media in

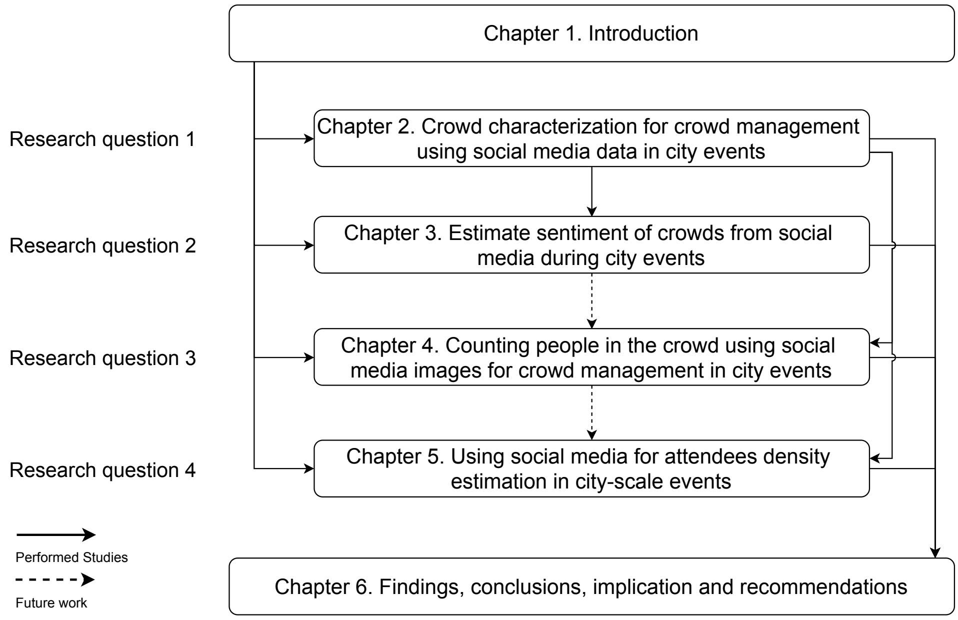


Figure 1.2: The schematic overview of the thesis.

city events. The method is based on elements of pedestrian traffic flow theory that were successfully assessed during city-scale events. The proposed method is validated in a case study and the result is compared with sensor data. The result of this chapter answers the research question RQ4.

Finally, Chapter 6 presents the findings and conclusions of this thesis, the implication for practice, as well as the recommendation for future research.

In addition, it should be noted that this thesis is a collection of articles. Chapter 3 and 5 correspond to articles that have already been published in scientific journals, and Chapter 2 and 4 correspond to articles that have been submitted for publication. Therefore, the chapters contain some repetitions in terms of information, particularly in the abstracts and introductory sections. The reader may want to skip these sections.

Chapter 2

Crowd Characterization for Crowd Management using Social Media Data in City Events

In this chapter, we characterise city events in terms of various aspects using social media data. This answers the first research question, i.e. **RQ1. To what extent are social media data able to characterize crowds in city events, in terms of demographic composition, city-role composition, spatio-temporal distribution, Points of Interest preferences and word use?**

To this end, we screen a set of factors (i.e. visitor profile, crowd size, density, mobility, location, and semantics) that characterize crowd behaviour and introduce a set of proxies (i.e. demographics, city-role, crowd temporal distribution, post position, Points of Interests, and word use) derived from social media data. Furthermore, we characterize the crowd in two city-scale events, Sail 2015 and King's Day 2016, in terms of these proxies, and comparing them with information collected from events organizers and programs.

Our findings show that it is possible to characterize crowds in city-scale events using social media data, thus paving the way for new real-time and planning applications on crowd monitoring and management for city-scale events.

This chapter is published as a journal article: Gong, V. X., Daamen, W., Bozzon, A., & Hoogendoorn, S. P. (2020). Crowd characterization for crowd management using social media data in city events. *Travel Behaviour and Society*, 20, 192-212.

2.1 Introduction

As cities compete for global importance and influence, city-scale public events are becoming an important ingredient to foster tourism and economic growth. Sports events, thematic exhibitions, and national celebrations are examples of city-scale events that take place in vast urban areas, and attract large amounts of participants within short time spans. The scale and intensity of these happenings demand technological solutions supporting stakeholders (e.g. event organizers, public and safety authorities, attendees) to monitor and manage the crowd.

These stakeholders aim to minimize the risk of incidents due to issues caused by external and internal threats. They normally apply predefined measures according to the qualitative interpretation of the crowd by policemen, stewards, or event organization employees.

As the efficiency and effectiveness of crowd management measures depend on pedestrian behavior (Still, 2000; Zomer et al., 2015), it is valuable for stakeholders to have information about the expected, and preferably actual, pedestrian behavior of the crowd. Pedestrian behavior is influenced by factors such as age, gender, and ethnicity (Martin, 2006). Insights into the distribution of these factors in an event's population can help estimating and predicting crowd behavior, and as such could be beneficial to crowd management.

However, information about these factors is difficult to capture. Traditionally, this information is manually sampled by stewards or staff members (Earl et al., 2004), a practice that is expensive and prone to biases. ICT solutions based on sensors (e.g. GPS, custom mobile apps) could provide spatio-temporal information (i.e. GPS coordinates and timestamps) that is useful to study crowd behavior (Jamil et al., 2015), but they are not broadly adopted, and might not provide demographic information. Camera sensors provide images or video clips which could be used to extract crowd features (Favaretto et al., 2016; Ryan et al., 2009), and detect crowd behavior (Wang et al., 2012a; Zhan et al., 2006) through image recognition techniques. However, accessing the images or video recordings of public area is computationally intensive, and often restricted due to privacy issues.

The advent of web-based technologies provides new social data sources that could be used to analyze and understand pedestrian behavior. Several platforms, such as Twitter, Instagram and Foursquare, are widely used. Social media content (e.g. text messages, images) is time-stamped and often geo-tagged, and it inherently containssni rich semantic information that could be used for characterizing the crowd from a pedestrian behavior perspective. For instance, the text content of posts sent by the crowd may indicate what the people are talking about, in order to see e.g. whether participants are enthusiastic about the event they are participating in or whether (security) issues are discussed. Likewise, the profiles of social media users can help to determine the crowd demographic characteristics. The rich semantic information makes social media a promising data source to provide information for crowd characterization in the city-scale events.

Previous works explored social media as data source to analyze various aspects about human behaviour and their characteristics for crowd management in context of city events. With regard to human travel behaviours, Rashidi et al. (2017) explored the capacity of social media data for modelling travel behaviour. Tyshchuk & Wallace (2018) explored a set of behaviors which are associated with warning response process using social media. Roy et al. (2019) quantified and analysed human mobility resilience to extreme events using geo-located social media. Krueger et al. (2019) proposed a visual analysis framework of city dynamics, including temporal patterns of visited places and citizens' mobility routines, using geo-located social media data. To explore the characteristics of human behavior, Abbasi et al. (2015) investigated a set of travel attributes which are extracted from social media data, such as trip purpose and activity location. Also several studies are performed in the context of city events. Yang et al. (2019a); Gao (2015); Hawelka et al. (2014); Yang et al. (2019b) use social media data collected in city events to investigate mobility issues. Cottrill et al. (2017) studied how attendees' behaviour are affected in a large city event, in terms of providing and sharing transport-related information and responding to requests, based on social media. Pramanik et al. (2019); Hochmair et al. (2018); Balduini et al. (2014) proposed methods to provide real-time Point of Interest (PoI) recommendations in city events using social media. Alkhatib et al. (2019) proposed a framework for monitoring incidents during events in cities. Though the utility of social media data has been shown in urban application domains, no previous work aimed at characterizing the crowd of city-scale events, with a specific focus on crowd management. What is lacking is an in-depth understanding of which factors could be extracted from social media data, and which automatic user modelling techniques can provide an accurate and reliable estimation of such factors.

In this paper, we perform a study to show to what extent social media data could be used for characterizing crowds in city-scale events using factors for crowd management. First, we identify a set of factors that are relevant for pedestrian behavior analysis for crowd management, and explore existing methods for extracting information about these factors from social media data. To showcase the application of these methods we perform two case studies having different properties. In each case, we collect social media data from multiple platforms, and extract the required information using SocialGlass (Bocconi et al., 2015), an integrated system for processing social media data. We then perform an exploratory analysis about these factors and correlate them with the corresponding event to check their accuracy and reliability. Discussions and conclusions are included at the end of the paper.

2.2 Crowd Characterization

In this work, we seek a better understanding of how social media data can be used to support crowd management. To this end, we provide insights about factors that are known to influence pedestrian behavior. In this section, we first introduce a selection of factors that are relevant to pedestrian behavior analysis; then we describe how such factors could be calculated from social media data.

Table 2.1: Influencing factors with corresponding social media proxies

Category	Factors	Social media proxy					
		Demographic	City-role	Crowd Temp. Dist.	Post position	PoI	Word use
Individual Characteristics	Demographic	x					
	Route familiarity		x				
	Perception of danger						x
	Type of destination					x	
Social Network	Household				x	x	
	Acquaintances				x		
	Neighborhood				x	x	
Trip characteristics	Trip purpose					x	
	Crowdedness						x
	Distance / proximity				x		
	Capacity			x			
	Traffic volume			x			
Built environment	Type of area					x	
	Percentage of foreigners		x				
	Aesthetics						x
	Distance to nearest transit stop				x	x	
	Population density			x			
	Intersection density			x			
	Road density			x			

Crowd Temp. Dist.: Crowd temporal distribution.

Table 2.2: An example of questions in crowd management plan and the social media proxies which can help answering these questions

Questions	Social media proxies
What is the demographics composition of the participants?	Demographics
What is the percentage of people from other cities?	City role
What is the crowd density during the event?	Crowd temporal distribution
Where is the most crowded area?	Post position
What kinds of places are to be most visited by the crowd in different region?	PoI
What is the sentiment of the crowd?	Word use

2.2.1 Characterization factors

Following the above discussion, criteria for selecting factors are:

(1). The factors should be identified as influencing the pedestrian behavior.

(2). The factors should be derived from social media.

As mentioned before, a set of factors has been discussed in (Martin, 2006) that affect pedestrian behavior. These factors can be classified into 6 categories, being Individual characteristics, Social network, Trip characteristics, Built environment, Destination environment and Physical environment. These factors with the corresponding social media proxies are listed in Table 2.1. These factors influence different types of pedestrian behavior, i.e. activity choice behavior, destination choice behavior, mode choice behavior, and route choice behavior (see (Wegener, 2004; Hoogendoorn & Bovy, 2005; Daamen, 2004) and Figure 4.1). Obtaining information about these factors may help with understanding such types of pedestrian behavior and further support crowd management.

As indicated in the previous section, crowd managers usually apply predefined measures according to the information about these factors. This is implemented in a crowd management plan (Still, 2014; Tubbs & Meacham, 2007; Abbott & Geddie, 2000), in which a set of questions are to be answered. Answering these questions require qualitative and quantitative information about the crowd. Examples of these questions or required information in a crowd management plan are shown in Table 2.2.

In the following sections, we explain why those factors and the social media proxies are connected, and which methods are used to calculate these proxies from social media data.

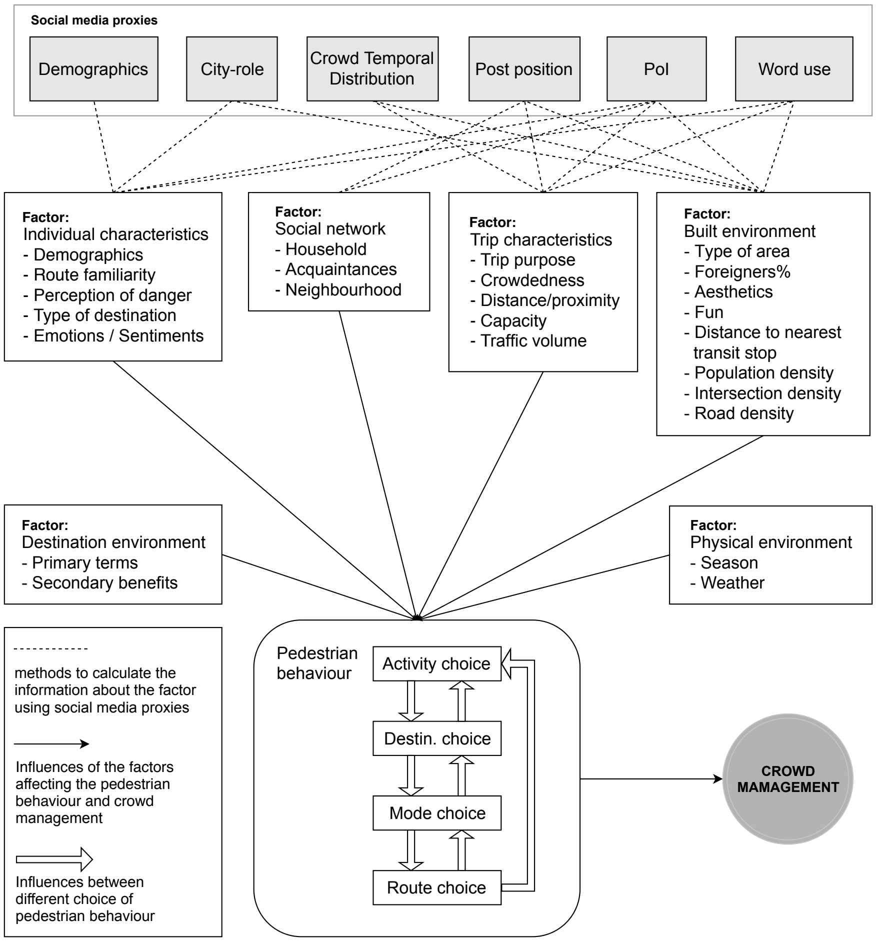


Figure 2.1: Illustration of relationship between crowd management, pedestrian behavior, influencing factors, and social media proxies. The numbers in the brackets denote the references.

2.2.2 Social Media Data Analysis for Crowd Characterization

Among all popular social media platforms, we focus our studies on three applications that provide data retrieval APIs, namely Twitter, Instagram and Foursquare. Twitter and Instagram provide posts and user profiles, while Foursquare provides Points of Interest (PoI) – the information about a location where people send posts. Twitter is a text-based social media platform, and one of the oldest social networking applications; Instagram is an image-based social media platform, which is particularly welcomed by female users (Yang et al., 2016; Gong, 2016). Data available for retrieval from such platforms include user profile information and submitted posts, their content, and time-stamp. A certain amount of posts contain coordinates where posts are sent, namely the geo-referenced posts, and the PoI information determined from geo-referenced posts. As city events take place at specific locations or areas, in this study we focus on social media data with geo-referenced posts. Based on the collected social media data, several proxies for crowd characterization factors can be calculated, namely demographic characteristics (i.e. age, gender), city-role, post spatio-temporal distribution, PoI, and word use. Each of them is addressed in the following subsections.

These proxies are used in crowd management following a two-phase approach (Martella et al., 2017), i.e. the planning phase and operational phase. In the planning phase, crowd managers require the past event data to infer guidelines and to perform computer simulations of the crowds in the event. In the operational phase, a set of what-if scenarios are proposed. Crowd managers require the current information of the crowd to decide which scenario is best to predict and prevent accidents. Further, they can apply feasible predefined measures in the given scenario. For example, having the profile information about the event visitors, such as age and gender distributions, crowd managers can formulate event guidelines suitable for these visitors in the planning phase, and prepare a set of what-if scenarios, such as what to do if most visitors are young people and the density of visitors reaches a certain threshold.

The accuracy of techniques to derive these proxies is influenced by the amount and representativeness of information such as user profiles. Though social media is not used by everyone in the event, it could be considered as a partially representative sample of the crowd during events.

Demographic characteristics

Demographic characteristics, i.e. age(Berrigan & Troiano, 2002), gender (Berrigan & Troiano, 2002; Panter & Jones, 2010), have been identified as factors affecting pedestrian activity choice, destination choice, mode choice and route choice. This information could be derived from social media by approaches using text categorization (Peersman et al., 2011), first name (Lansley & Longley, 2016; Mislove et al., 2011), and profile picture (Bocconi et al., 2015; Longley et al., 2015). In our study, we use the user profile picture to determine user's age (Bocconi et al., 2015; Zhou et al., 2015; Psyllidis et al., 2015a), and a multi-modal decision tree classifier (Yang et al., 2016; Titos Bolivar, 2014) combining the user's profile picture (Zhou et al., 2015) and the first name (Lansley & Longley, 2016) to detect user's gender information. A manual check with 628 labelled social media profiles (Yang et al., 2016) shows that both age and gender detection reach promise performance, i.e. 88% precision for age detection when faces are present, and 85% precision for gender detection.

People can be classified according to different indicators. One of the well-known ones is gender, i.e. male and female. Age is also known to influence behavior, often using four groups (Berrigan & Troiano, 2002). The range of each age group is defined considering social and physiological science (Al-Zahrani et al., 2003; Young et al., 1993) as follows:

- Young: user between 0 and 18,
- Young-adult: user between 18 and 30,
- Adult: user between 30 and 65,
- Old: user older than 65.

City-role

The city-role describes the relationship between the people and the city, i.e.:

- Resident: attendees living in the city of the event;
- Local traveler: attendees living in the same country, but in another city;
- Foreign traveler: attendees from a foreign country.

The percentage of foreigners (Kim et al., 2014; Rietveld & Daniel, 2004) and people's familiarity with a route (Kim et al., 2014) are identified as factors affecting mode choice behavior (Kim et al., 2014; Rietveld & Daniel, 2004) and route choice behavior (Kim et al., 2014), respectively. Information about these factors can be derived using social media by checking a user's home location through a recursion search method (Cheng et al., 2011; Titos Bolivar, 2014), which shows promise accuracy (covering about 0.004 square miles) according to the comparison (Yang et al., 2016).

Crowd temporal distribution

The temporal distribution of a crowd, i.e. the distribution of persons present at a certain area over time, is identified as a factor affecting destination choice (Han et al., 2010; Zahran et al., 2008), mode choice (Handy, 1996; Zahran et al., 2008; Guo, 2009; Rodríguez et al., 2009) and route choice (Zahran et al., 2008; Guo, 2009), as illustrated in Figure 4.1. Calculating the temporal distribution of the crowd during an event requires information about the amount of people in an event area during a predefined period of time.

In social media, each post is sent with a timestamp. This information may be used to count the amount of posts sent by different people in a period of time. It is then used as a proxy for the temporal distribution of crowds (Yang et al., 2016; Gong, 2016; Titos Bolivar, 2014), which is temporally correlated with the estimated ground truth from sensor data according to a comparison (Gong et al., 2018b).

Post position

Distance/proximity (van der Waerden et al., 1998; Maley & Weinberger, 2011; Panter & Jones, 2010) is identified as a factor affecting all four pedestrian behaviors mentioned in Figure 4.1. To calculate the distance, e.g. the distance between a pedestrian and a certain object in the area, having the position of the pedestrian is required.

In social media, the geo-referenced posts contain the coordinate of the location they have been sent. This position data can be a proxy to calculate distances (Yang et al., 2016; Gong, 2016; Titos Bolivar, 2014).

Points of Interest

Factors such as type of destination (Eash, 1999), diversity of land use (Rodríguez et al., 2009; Panter & Jones, 2010; McCormack & Shiell, 2011), and trip purpose (Handy, 1996) are identified as factors affecting destination choice (Eash, 1999), mode choice (Eash, 1999; Handy, 1996), and route choice (Rodríguez et al., 2009; Panter & Jones, 2010; McCormack & Shiell, 2011), respectively. These factors require information about a location with its functionality category as well as popularity, which can be provided by the Point of Interest (PoI), a particular location that someone may find useful or interesting, such as a hotel, a restaurant, or a bus station. A social media post sending from a PoI indicates a PoI has been visited by this user. With such information, we may extract the set of PoIs visited by people during an event, as well as PoI functionality categories and popularity. The destination of a pedestrian's trip as well as the trip purpose could be examples for which the data can be analyzed.

The PoI information can be derived from social media through various techniques, such as Natural Language Processing (Lingad et al., 2013), user relationship analysis (Davis Jr et al., 2011), and the Venues Mapping method (Noulas et al., 2012). The Venues Mapping method proposed by Noulas et al. (2012) establishes a model to determine the venue visited by each user considering multiple aspects in their approach, i.e. popular places, similar places, users' preferences in selecting places, places visited by friends, and places in short distance. In our research, we employ the Venues Mapping method (Noulas et al., 2012) to get the PoI visited by social media users as it results in 5% to 18% improvement over other methods (Noulas et al., 2012). We record the top-level PoI category defined by Foursquare visited by social media users for analysis.

Word use

Influencing factors such as Crowdedness (Pratiwi et al., 2015; Duives et al., 2016), Aesthetics (Guo, 2009; Panter & Jones, 2010; McCormack & Shiell, 2011), Fun (Florez et al., 2014), and Perception of danger (Panter & Jones, 2010) affect mode choice and route choice. These factors require information about a pedestrian's expressions and feelings. This information can be derived from social media data.

A social media post usually consists of a texture attribute which can be used to infer topics the people talk about, and their feelings. In this research,

Table 2.3: The measurements of social media proxies to derive property of factors

Aspect	Proxy	Measurement
Demographics	Gender	#male, #female, M/F
	Age	#young, #young-adult, #adult, #old, SD
City-role	City-role	#resident, #local_traveler, #foreign_traveler, SD, R/L
Crowd temp. dist.	Post amount	#GP of day, Max #GP and Time, Min #GP and Time
Position	Coordinates	latitude, longitude
PoIs	PoIs	#PoI_visit
	PoI category	#PoI functionality category
Word use	Text content	Word count

#male: number of people determined as male,
Crowd temp. dist.: Crowd temporal distribution.
M/F: the rate of Male with Female,
SD: Standard Deviation.
R/L: the rate of Resident with Local.
#GP: amount of Geo-posts.
Max #GP and Time: the max amount of Geo-posts, and the time of a period during which this amount is observed.

we visualize the frequently used words (word-cloud) sent by the crowd in order to provide such information, see also (Yang et al., 2016; Schwartz et al., 2013; Chen et al., 2014; Gong, 2016).

Summary

The sections above introduced a set of factors, about which information can be derived from social media data, the so-called proxies. We further described each proxy with properties and methods to calculate them. An overview of the measurements of the proxies is shown in Table 2.3. In the remainder of the chapter, we will apply these techniques in two city events and compare and analyze the estimated information with the events programs.

2.3 Applying crowd characterization based on Social Media data in two city-scale events

In this section, we showcase how social media data (and related methods) can be used to characterize the crowd in two city-scale events by providing information about the factors described in the previous section. Furthermore, we relate the derived information with the event programs, to discuss its accuracy and reliability.

2.3.1 Case selection

We investigate two events that took place in Amsterdam, the Netherlands, respectively Sail 2015 (in the following referred to as Sail) and King's Day 2016 (Kingsday). We selected the two events for their similarities and their differences. On the one hand, these events have similar properties, being *city-scale*, and taking place in the same *urban environment* and *planned*, *temporally constrained*, and *thoroughly organized* (in contrast to seasonal events, such as Christmas shopping, or serendipitous events, like protests) and *popular* and *generalist*, as they attract large crowds with diverse demographics.

On the other hand, the two events also differ from each other in terms of duration, topic, crowd composition and event terrain. For instance, for duration, Sail lasts for 5 days, ending in a weekend, whereas Kingsday is a single-day event, and a public holiday. As to the topic, Sail being a naval event offering, for instance tall-ship exhibition, nautical history experience, fireworks show, while Kingsday is a recurrent national celebration, which offers a boat parade, free market and parties. As for the crowd composition, Sail is known to attract visitors from the whole world, while Kingsday is a national event. For the event terrain, Sail has activities centred around the IJhaven area (where ships docked), while Kingsday activities are scattered throughout the city.

To compare the analysis with the actual situation where quantitative ground truth existed, we also perform an analysis for two terrains (sub-area), based on the findings in our previous work (Gong et al., 2018b), where the number of people calculated from social media in each terrain is temporally correlated with the estimated ground truth calculated using sensor data. The social media users in these two terrains, i.e. Javakade in Sail and Zuidplein in King's Day. The Javakade located on Java Island, directly faces the IJHaven, the bay area where the boats docked. This terrain is residential, with no recreational businesses. Areas separated by canals are connected by small pedestrian bridges, where several docked boats can be accessed during Sail event. The Zuidplein is the forecourt of the station Amsterdam Zuid, which is a popular pedestrian square connecting the station with the CBD area, and the Amsterdam OUD-Zuid. Around the square, there are vari-

Table 2.4: Number of users of which the demographic and city-role have been derived from social media in two terrains during the Sail and King's Day events, respectively.

		Terrain	Age					Gender			City-Role			
			Young	You. Adult	Adult	old	Sum	Male	Female	Sum	Resident	Loc. Tour	For. Tour	Sum
Sail	Twitter	Javakade	8.4%	39.6%	52.0%	0.0%	94	68.2%	31.8%	163	44.6%	33.6%	21.9%	187
	Instagram		19.4%	49.5%	31.0%	0.2%	367	46.5%	53.5%	757	48.8%	21.9%	29.2%	1018
King's Day	Twitter	Zuidplein	12.9%	43.9%	43.3%	0.0%	69	61.6%	38.4%	98	45.6%	15.2%	39.2%	191
	Instagram		23.3%	49.8%	26.9%	0.0%	637	39.6%	60.4%	1032	44.0%	21.1%	35.0%	3965

You. Adult: Young Adult.
Age for Young: 0-18, **Young Adult**: 18-30, **Adult**: 31-64, **Old**: 65+.
The scope of the terrains are illustrated in **Figure 3** in Chapter 5.
The users in each terrain is identified using **speed- and flow-based** density estimation methods (K3/K4) in **Table 2** in Chapter 5.

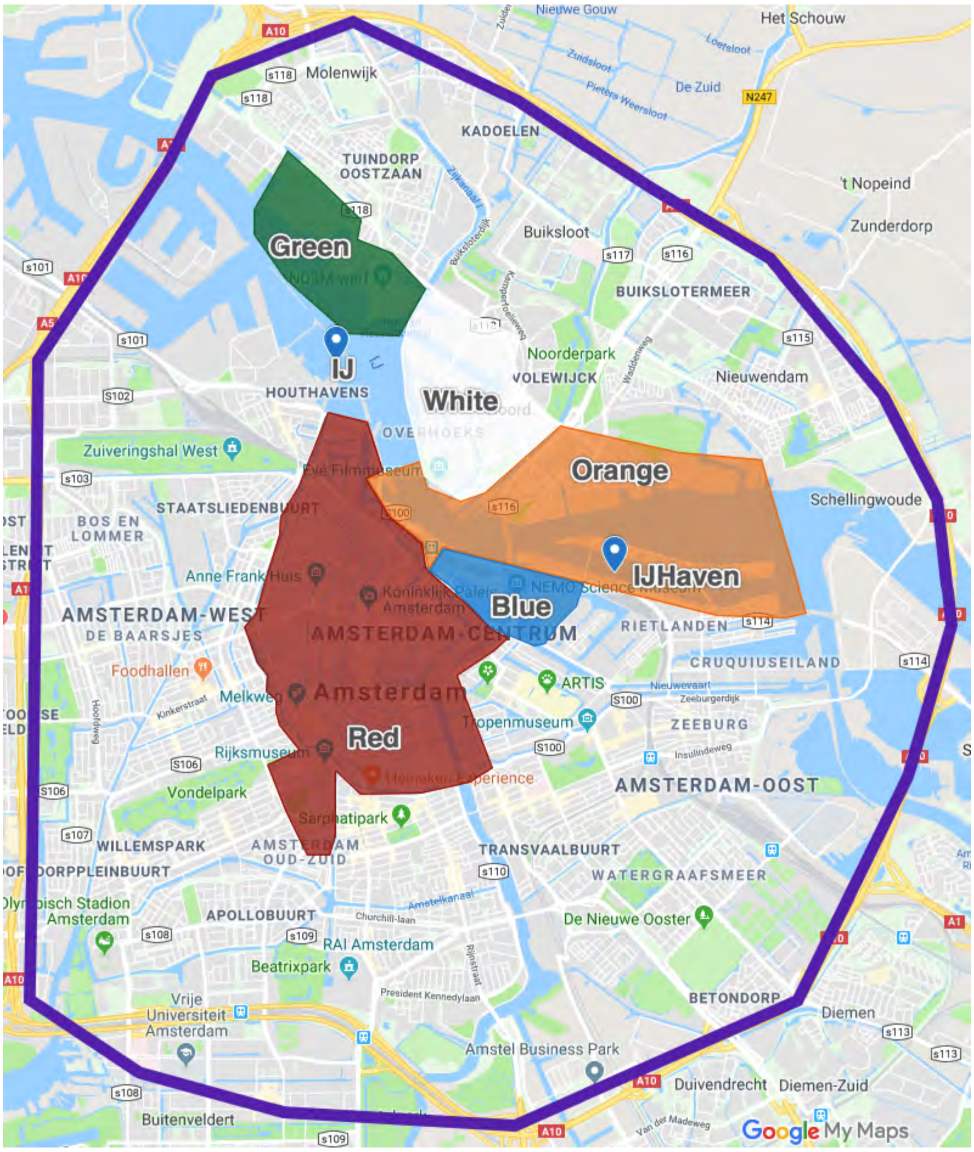
ous shops and restaurants, attracting a large amount of people during King's Day event. The number of users of which the demographic and city-role have been derived from these two terrains are listed in Table 2.4.

We selected these two events to compare the introduced crowd characterization for events with different fingerprints. The areas where these two events took place are shown in Figure 2.2. Further details about the events are introduced in the following sub-sections.

Case 1: Sail 2015

SAIL Amsterdam is a quinquennial maritime event in Amsterdam, the Netherlands. Tallships from all over the world come to the city to be visited and visitors join activities. It is the largest public event in the Netherlands: the 2015 edition of the event lasted 5 days, from August 19 to 23, and attracted more than 2 million people. The exhibition included tallships and historical ships, as well as a large number of other boats. The official event area covers most of the city center, and was organized into five so-called oceans, each devoted to a theme.

The program of this event included sub-events spanning all five days. On August 19, all tallships sailed from the coast towards Amsterdam and docked in the IJhaven. During the following three days, the tallships were open for visits from 10AM till 11PM. A set of ship related activities took place around IJhaven attracting a huge number of people who are interested in this topic. The ships departed again on August 23 in the closing SAIL-out event after a Sail Thank You parade. Every day, a firework show took place at IJhaven lasting for 15 minutes between 22:00 to 23:00.



(a) Area of two events that took place in Amsterdam. Activities during Kingsday took place in the whole city of Amsterdam (area bounded by dark blue line). The other 5 coloured areas are for Sail, i.e. Orange, White, Blue, Green and Red activity areas, the so-called Oceans. Marked locations are further explained in the case introduction and analysis.

(b) Sail 2015 event

(c) King's Day 2016 event

Figure 2.2: The two events in Amsterdam selected in this study

Case 2: King's Day 2016

King's Day is a national holiday held each year in April, celebrating the birthday of King Willem-Alexander. In major cities in the Netherlands it is celebrated with joyful open air festivities. People join this yearly regular event with their families and friends. In 2016, the King's Day celebration attracted more than 1.5 million people in Amsterdam, including Dutch tourists and a huge number of foreign tourists.

Though it is a one day public holiday, it is certainly not a day of rest. The celebrations started on the eve of King's Day - named as the King's Night. Parties, music, and carnival atmosphere continuing throughout the city until the end of the day. Following King's Night, the most interesting activity on King's Day in Amsterdam is the boat parade. From 1 pm, canals are packed with boat parties, during which the boats are sailing along the canals throughout the city with people enjoying drinking and celebrating wearing orange. Besides, several large museums are open for people who would like to experience culture and history.

2.3.2 Data collection

For each case, we collected geo-referenced social media data on the Twitter and Instagram platforms. The geo-referenced social media posts were mapped with PoIs from Foursquare. Then, we derived information about the crowd, including age, gender, city-role, crowd temporal distribution, post position, PoIs, and word use. We analyzed the derived information for each case, looking for meaningful relationships with the events' programs. We also compared the outcomes of the analysis of the two cases, highlighting similarities and differences.

The data is collected and derived using SocialGlass (Bocconi et al., 2015), an integrated system for crawling and processing social media data. First, we set up a crawling task with a duration (starting and ending date) and an event area (a bounding-box for Twitter, and multiple circles for Instagram) to crawl geo-referenced social media posts sent during an event, through queries on Twitter and Instagram. Second, we screen out unique users from the captured social media posts, as one user may send multiple posts. Third, we crawl user profile data on both platforms. Next, we crawl historical geo-referenced posts for each user on Twitter and Instagram, respectively. Further, we calculate the demographic, city-role and word use information for each user, and generating PoIs information through venue mapping algorithm. Finally, we export demographic, city-role, crowd temporal distribution, position, PoIs and word use information from the SocialGlass platform.

With regard to Sail 2015, the data set includes posts generated from August 19 until August 23. For King's Day 2016, we collected data from April 26 to April 28, to respectively cover celebration starting the night before (i.e. King's Night), King's Day itself, and the following day (to capture celebrations lasting throughout the night).

In the crowd temporal distribution analysis, we further include social media data sent seven days before each event and seven days after, in order to compare the pattern of the crowd distribution between event days and regular days, as well as to compare event days with regular week and weekend days. However, for the other analyses we only use event dates, i.e. August 19 to 23 for Sail and April 26 to 28 for King's Day, shown in Table 2.5.

From both events we collected more posts on Instagram than on Twitter; it is caused by, on the one hand, the scarcity of the geo-referenced tweets which only accounts for 1-2% of all tweets (Paule et al., 2019). We decide to obtain more geo-referenced tweets in our future work using techniques such as geolocalisation (Middleton et al., 2018; Paule et al., 2019). On the other hand, it suggests that Instagram, being image based social media, is a preferred social networking choice during an event or festival. This result is consistent with findings from related work (Yang et al., 2016; Gong, 2016; Titos Bolivar, 2014). The sparsity of social media data currently may affect the representativeness of information for crowd management when deriving information from geo-referenced Tweets. However, such influence may be reduced by increasing the collecting of amount of social media data and applying geolocalisation methods.

The sparsity of social media data currently may affect the representativeness of information for crowd management. However, such influence may be reduced by increasing the amount of social media data collection and applying applying geolocalisation methods. Results also show that the amount of data collected during Sail is more than during King's Day except for the number of users on Instagram (Sail, Instagram users: 27.3k; King's Day, Instagram users: 28.5k), which indicates that Instagram has been used by more people during King's Day than during Sail.

The number of users we collected on Twitter and Instagram may contain duplicated users, i.e. those who have profiles and use these two social networks in the same event. This is indeed a bias in the dataset. To counter

Table 2.5: Overview of the data collected during Sail 2015 and King's Day 2016 in Amsterdam. The number of Users, Geo-Posts, and PoIs are expressed in thousands.

	Sail 2015 Aug 19-23, 2015		King's Day 2016 Apr 26-28 2016	
	Twitter	Instagram	Twitter	Instagram
#User	2.8	27.3	1.6	28.5
#Geo-Posts	11.6	60.4	4.6	44.1
#PoIs	2.5	8.0	1.4	2.4

this bias, we perform a manual check on counting the number of users who mention their tweets on Instagram posts, or mention their Instagram posts on tweets. The results show that less than 1% of users performed this operation. For users whose Twitter and Instagram accounts do not show any explicit connections, we may not be able to count their times of appearance. Even though, the number of such users may not be high. We therefore assume that in this research all social media users collected on Twitter and Instagram are individual users.

2.4 Findings & Analysis

In this section, we analyze the collected social media data on both social media networks, i.e. Twitter and Instagram, in terms of demographics, city-role, crowd temporal distribution, post location, PoIs and word-use. For each aspect, we compare findings across different events and social media platforms, and relate them to the respective event programs.

2.4.1 Demographic analysis

The demographic analysis consists of the analyses of age and gender.

Age. Table 2.6 shows that more young-adult users (Sail: 42.7%, 49.8%; King's Day: 44.1%, 44.6%) are captured in both social media across events, followed by adult users and young users. Old users are extremely sparse, a result that we attribute to distinct technology penetration – old people seldom use social media. In the meantime, Instagram is far more used by young-

Table 2.6: Number of users of which the demographic and city-role have been derived from social media during the Sail and King's Day events.

		Age					Gender			City-Role			
		Young	You. Adult	Adult	Old	Sum	Male	Female	Sum	Resident	Loc. Tour	For. Tour	Sum
Sail	Twitter	12.3%	42.7%	45.0%	0.0%	1225	56.7%	43.3%	2267	37.8%	21.3%	40.8%	2779
	Instagram	24.4%	49.8%	25.8%	0.1%	7130	42.0%	58.0%	19341	35.3%	14.7%	50.0%	26947
King's Day	Twitter	13.4%	44.1%	42.5%	0.0%	515	56.8%	43.2%	754	36.3%	14.1%	49.6%	1640
	Instagram	23.0%	44.6%	32.3%	1.0%	8747	42.1%	57.9%	8510	40.1%	14.6%	45.4%	7028

You. Adult: Young Adult. **Loc. Tour**: Local Tour. **For. Tour**: Foreign Tourist.
Age for Young: 0-18, **Young Adult**: 18-30, **Adult**: 31-64, **Old**: 65+.

Table 2.7: Number of users of which the demographic and city-role have been derived from social media during three sub-events of Sail.

		Age					Gender		
		Young	You. Adult	Adult	Old	Sum	Male	Female	Sum
Sail fireworks	Twitter	9.1%	42.7%	48.2%	0.0%	110	61.0%	39.0%	182
	Instagram	21.3%	50.5%	27.1%	1.1%	727	44.6%	55.4%	1403
Sail topic activities	Twitter	8.9%	38.3%	52.8%	0.0%	358	65.0%	35.0%	609
	Instagram	21.0%	47.7%	31.2%	0.1%	1745	46.1%	53.9%	3645
Sail parade	Twitter	6.3%	38.5%	55.2%	0.0%	96	59.8%	40.2%	132
	Instagram	20.3%	48.8%	30.9%	0.0%	602	43.1%	56.9%	954

You. Adult: Young Adult.
Age for Young: 0-18, **Young Adult**: 18-30, **Adult**: 31-64, **Old**: 65+.

adults, which might be because people in this age tend to share pictures taken during enjoyable events. The standard deviation of age of people during King's Day is lower than during Sail, indicating that more people across age ranges make use of social media during King's Day. The percentages of young and young-adult people observed during King's Day (Twitter: 13.4%, 44.1%, Instagram: 23.0%, 44.6%) are almost the same as during the Sail event (Twitter: 12.3%, 42.7%, Instagram: 24.4%, 49.8%).

We further zoom into 3 sub-events which attracted attendees with different demographic characteristics during Sail. According to the Sail official website, everyday between 09h and 21h – except for the last day – naval-related activities were organized around the IJhaven area where boats docked. Examples of such activities, which attracted people interested in this field, were tall-ship exhibitions, nautical history experiences, sports games on the water, and first aid in boat damage training. In contrast, the fireworks shows took place every night between 22h and 23h, and attracted more families [1]. The Sail parade had similar population distribution, but took place during the last day of the event between 12h and 18h. Table 2.7 shows the number of people in different age and gender groups of this three sub-events. The standard deviation of age during the Sail parade (Twitter: 23.9, Instagram: 86.9) is less than during the Sail fireworks show (Twitter: 23.3, Instagram: 112.3), followed by the Sail topic related activities (Twitter: 80.0, Instagram: 235.7). It is in accordance with the expectation that the Sail fireworks and parade attract more families which are more evenly distributed populations.

With regard to the social media users in two terrains according to Table 2.4, the standard deviation of age in Javakade in Sail event (Twitter: 56.2, Instagram: 267.2) is larger than in Zuidplein during King's Day event (Twitter: 37.7, Instagram: 45.4). This is in line with the fact of different types of PoIs in terrains; Recreation amenities, such as bars, clubs, shops and restaurants which located in Zuidplein may attract more young people thus lead to more social media posts sent by young and young-adult than in residential area.

Gender According to Table 2.6 in both events, more male users are detected on Twitter while more female users are on Instagram, indicating that sharing pictures in such festivals is more popular among female users. The ratio between male and female users is similar for the two events (Twitter:

[1] https://www.abroad-experience.com/blog/sail-amsterdam-maritime-fun-for-the-whole-family/

1.3, Instagram: 0.7). With regard to the three sub-events shown in Table 2.7, the standard deviation of ratio between male and female (Twitter: 61.4, Instagram: 57.3) is less obvious compared with standard deviation in age (Twitter: 32.6, Instagram: 79.5).

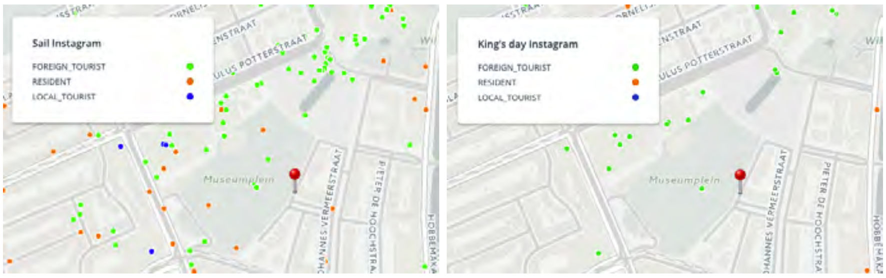
With regard to the gender distribution in the two terrains shown in Table 2.4, the standard deviation of gender on Instagram (Javakade: 131.5, Zuidplein: 53.1) is larger than on Twitter (Javakade: 101.8, Zuidplein: 39.6). In particular on Instagram, the gender of people in Javakade in Sail event is more equally distributed than on Zuidplein in King's Day. This may be caused that more recreation and activities on Zuidplein during King's Day gives rise to the usage of Instagram, the image based social media network, which are more popular for female users to share their feelings (Yang et al., 2016).

City-role

Table 2.6 shows that more foreign travelers than residents and local tourists are identified on social media in both events. As a popular touristic city, Amsterdam usually attracts huge numbers of tourists, and large scale events only increase these numbers. It is also likely that foreign travelers are more active on social media as they feel fresh and excited to be in a new place. This is also observed in related work (Yang et al., 2016). The ratio between local travelers and residents during King's Day is lower than during Sail, thus showing that Sail, an event taking place every 5-years, is more likely to attract visitors from other cities and countries. According to Table 2.5 and Table 2.6, the proportion of Kingsday Instagram users whose city-role is detected only accounts around 24.67% among all Instagram users who sent geo-reference posts on Kingsday in the event area. The lower rate of city-role detection may be caused by the updated privacy protection settings on Instagram and the limitation of Instagram API.

With regard to the city-role on the two terrains, according to Table 2.4, more residents are observed than any other types of users on both social media networks. It is in line with the actual situation that either the Javakade or Zuidplein is not the tourist attractor during events. Consequently, less local and foreign tourists than residents are captured on social media.

We further zoom into the Museumplein in Amsterdam, a popular tourist attraction, to show how social media characterize the city-role in this area, as compared to events. The Museumplein is a famous tourist attraction area



(a) Sail, Instagram (b) Kingsday, Instagram

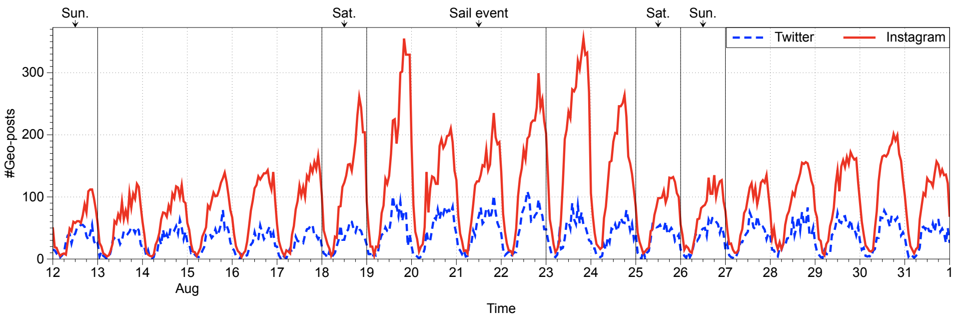
Figure 2.3: The city-role distribution around the Museumplein area from Instagram across events.

that hosts several popular museums (e.g. the Van Gogh museum, and the Rijksmuseum). Also, the square hosts the famous 'I Amsterdam Sign' attracting numerous tourists to take pictures with. As the Twitter data is too sparse for such a small area (only 13 tweets during Sail, and 8 during King's Day), we focus on Instagram data. According to Figure 2.3, more foreign tourists are observed in this area than local travellers and residents. This observation is in line with the fact that we see more pictures about this area on Instagram.

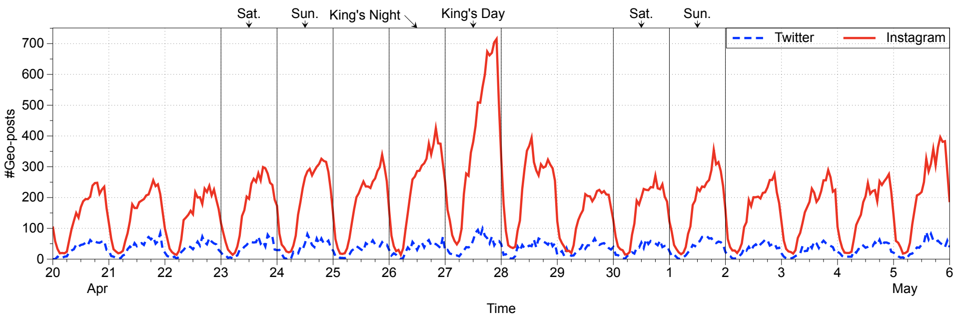
Crowd temporal distribution

Figure 2.4 shows the temporal distribution of social media activities observed around the two events in Amsterdam. In order to compare the pattern of the crowd distribution between event days and regular days, we show the temporal distribution during seven days before and seven days after the respective events. For each event the temporal distribution clearly illustrates the daily pattern. According to Figure 2.4, the amount of social media activities on event days increases faster than on regular days, and the total amount from one day before the event to one day after the event is larger than for ordinary days, including weekend days.

We further zoom into four attractive sub-events during Sail and King's Day event which attracts a large amount of people and may give rise to social media usage, listed in Table 2.8. We show the temporal distribution of social media activities per 30 minutes around each sub-event and compare it with three days before and three days after the event in the same time and area, shown in Figure 2.5 based on Instagram posts which are far more frequently



(a) Sail



(b) King' s Day

Figure 2.4: The temporal distribution of posts sent by people observed from social media.

Table 2.8: Sub-events of Sail and King's Day for temporal distribution analysis.

Sub-event		Date	Duration Time	# Length	Area in Amsterdam
Sail	SAIL-in Parade	19-08-2015	13:30 - 16:00, Day	3.5 hours	IJ, IJhaven
	Fireworks	19-08-2015 to 22-08-2015	22:00 - 23:00, Night	15 mins in 1 hour	IJhaven
King's Day	King's Night	26-04-2016 to 27-04-2016	18:00 - 02:00, Night	8 hours	City Centre
	King's Day Boat Parade	27-04-2016	13:00 - 17:00, Day	4 hours	City Centre

observed than Twitter.

We found that during the Sail-In parade and the King's Day boat parade the temporal distribution of social media activity shows significantly distinct patterns compared with three days before and three days after the event, which is in line with the reality that such sub-events attract a large number of people. During the other two sub-events, i.e. Sail Fireworks show and King's Night celebration, the temporal distribution of social media activity shows less distinct patterns compared with three days before and three days after the event. This observation indicates that sub-events such as the boat parade, which lasts for several hours and takes place in a city-scale area with magnificent views over the day is suitable to be characterized through a temporal distribution based on social media. Details about temporal distribution analyses of social media activities for the two events are presented in the Appendix.

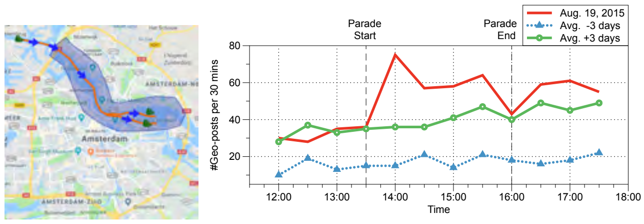
Post position

The positions of crowds in a city-scale event can be derived from social media data, as described in Section 2. Figure 2.6 shows the position heatmap of social media users observed during the two events in Amsterdam. It shows that during both events crowds are active on social media mostly in the city center. However, attendees of Sail were more active in the area around the IJ and IJhaven, a clear indication for crowd managers that special activities taking place around that area during Sail event which attract a large amount of people.

We could also observe event-related social media activities outside the event area in Figure 2.6, which may be caused by people who sent posts after attending the event.

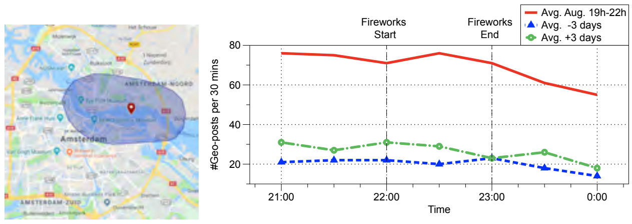
PoI

As illustrated in Section 2, the PoI data help with deriving information from pedestrian behavior factors such as built environment, trip characteristics and social networks. We first zoom into the area of Central Station, which is the major transportation hub in Amsterdam. According to Figure 2.7 (a) and (b), there are more Travel & Transport PoIs (red dot) visited by social media users than other PoIs, which corresponds to the expected indication



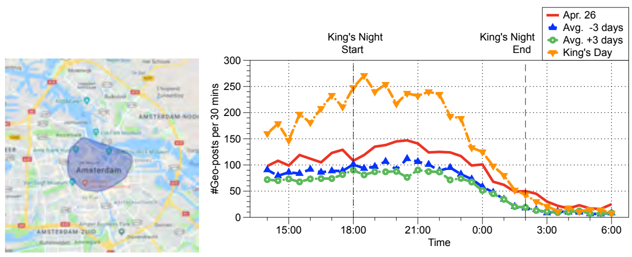
(a) Sail, In_Parade area

(b) Sail, In_Parade



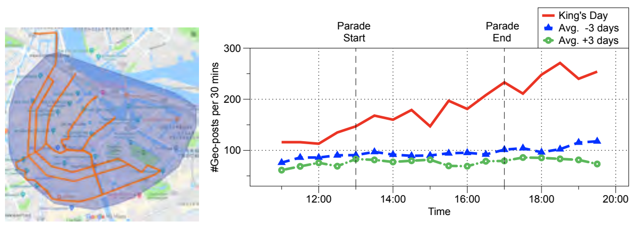
(c) Sail, Fireworks monitoring area

(d) Sail, Fireworks



(e) King's Day, King's Night area

(f) King's Day, King's Night

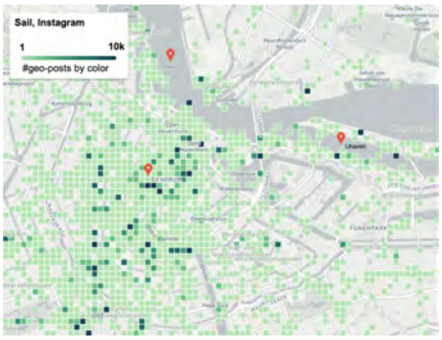


(g) King's Day, Boat Parade area

(h) King's Day, Boat Parade

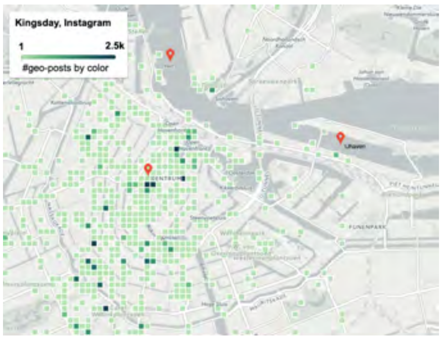
Figure 2.5: Temporal distribution of social media activity on sub-events of Sail and King's Day.

(a) Sail, Twitter



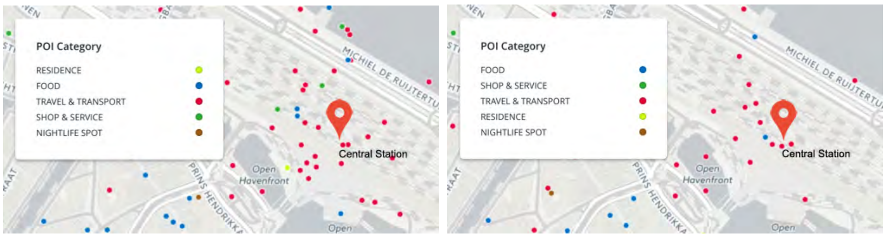
(b) Sail, Instagram

(c) King's Day, Twitter



(d) King's Day, Instagram

Figure 2.6: The heatmap of the position of people by color observed from social media posts.



(a) Sail, Central Station (b) King's Day, Central Station

(c) Sail, PoI population (d) King's Day, PoI population

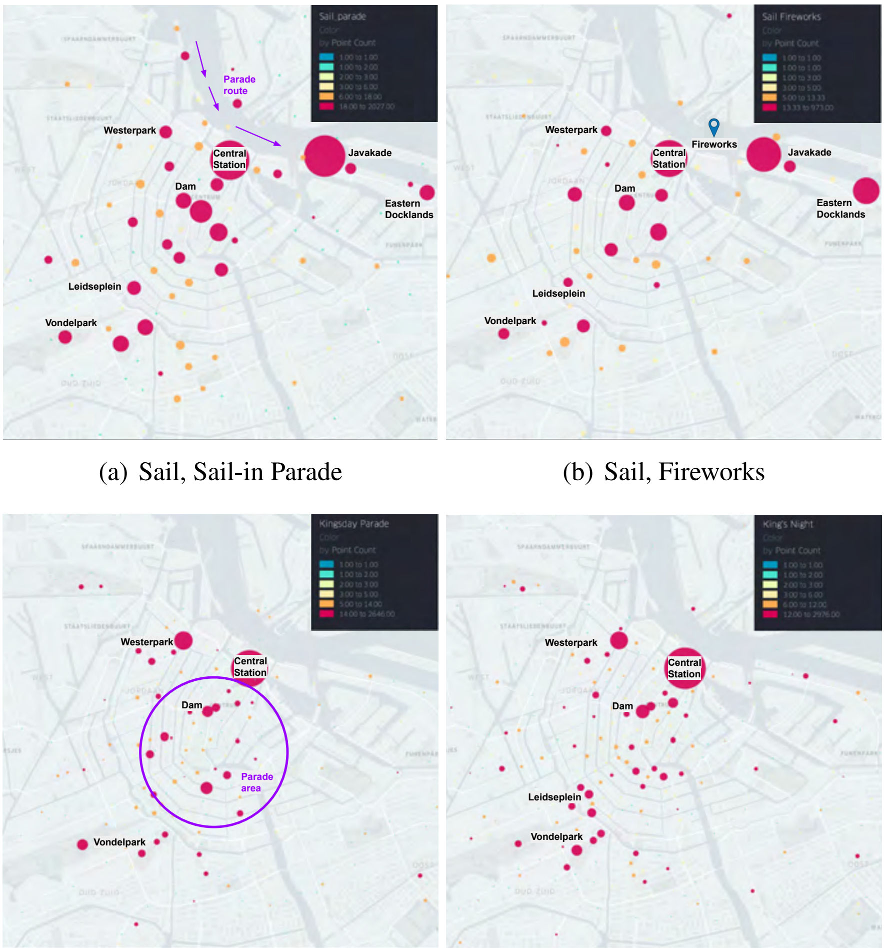
Figure 2.7: The PoIs observed from social media across events.

of the land use of this area according to the land-use of Amsterdam [2]. According to Figure 2.7 (c) and (d), popular PoIs during Sail are spread around 5 oceans, particularly in the Orange ocean (IJhaven area) where the boats docked, while during King's Day the popular PoIs are less distributed in the IJhaven area.

We can also use social media posts to discover PoIs which are most visited in different events. By clustering these PoIs we can further detect the most popular area visited by people during events, i.e. the Area of Interest, which is valuable for crowd managers to understand the most important area in different events.

To showcase the Area of Interest discovered using social media for crowd management, we generate a clustered map for each sub-events in 4 sub-events (Table 2.8) using DBSCAN algorithm (Ester et al., 1996) which groups together points that are closely packed together. According to Figure 2.8 (a)-(d), Amsterdam Central Station is a popular area across sub-events, which is natural as it is the largest transport hub in the city centre. When compared

[2]https://maps.amsterdam.nl/grondgebruik/?LANG=en



(a) Sail, Sail-in Parade (b) Sail, Fireworks

(c) King's Day, Boat Parade (d) King's Day, King's Night

Figure 2.8: The Area of Interest detected by clustering PoIs using social media posts for 4 sub-events listed in Table 8. The size and colour of each plot indicate the number of visits of the covering area. The bigger size in Red colour denotes larger popularity, while smaller size in Blue colour denotes less popularity.

with other transport hubs in the city such as Amsterdam Zuid, Amsterdam Central Station is more relevant for events especially for Sail, as it is closer to the event side. During both Sail sub-events, shown in Figure 2.8 (a) and (b), area close to IJ (bay area), such as Javakade and Eastern Docklands, are more popular than in King's Day event, which is in line with the fact that most activities during the Sail event are carried out around IJ. In particular, during the Fireworks show in the night during the Sail event, shown in Figure 2.8 (b), the Eastern Docklands are getting more popular. It may indicate that Eastern Docklands is a good place to have a full view of fireworks and thus attracts a large number of people. While, during King's Day, according to Figure 2.8 (c) and (d), the most popular areas are distributed in the city center rather than the IJ area. During the Kingsday boat parade, shown in Figure 2.8 (c), the most popular areas include the city center where the parade is taking place, the Westerpark where music performances are being held, and the Vondelpark where activities such as flea markets are carried out. During King's Night, according to Figure 2.8 (d), the most popular area, besides the Amsterdam city center, are the Westerpark and Leidseplein, where music performances and clubs (or bars) attract a large number of visitors, respectively. Crowd managers may perform measures in these areas to avoid crowdedness during sub-events. Details about clustering analyses of social media activities for 4 sub-events are presented in the Appendix section.

Word use

With regard to the word use of the crowd, we generate word clouds showing the most frequently used words observed from social media. The word clouds are generated using text content of the posts filtering out hashtags and URLs from Twitter and Instagram, respectively. As shown in Figure 2.9 (a)–(d), the general words about the city (e.g. Amsterdam, Netherlands) and event name related (e.g. Sail: SAIL. King's Day: King's, King's Day, King's night) in either English or Dutch are most frequently used across events. In order to characterize what people talk about discarding these words, we exclude them and re-generate word-clouds in Figure 2.9 (e)–(h). Results shows that activity related words (e.g. Sail: ship, drinking, parade. King's Day: drinking, orange, kingsland) and emotional words including emojis are captured (e.g. happy, love, vertraagd (delayed), vechtpartij (fighting), best, amazing, exciting, fun, lovely). It indicates that the popular topics and sentiment of people in the crowd which are valuable for crowd management can

(a) Sail, Twitter
(b) Sail, Instagram
(c) King's Day, Twitter
(d) King's Day, Instagram
(e) Sail, Twitter, filtered
(f) Sail, Instagram, filtered
(g) King's Day, Twitter, filtered
(h) King's Day, Instagram, filtered

Figure 2.9: Word-clouds during Sail and King's Day.

(a) Javakade, Sail, Instagram, filtered (b) Zuidplein, King's Day, Instagram, filtered

Figure 2.10: Word-clouds in Javakade during Sail, and Zuidplein during King's Day.

Table 2.9: Word count of 'crowded' in word use across events.

Event		Twitter				Instagram			
		Crowded		Druk		Crowded		Druk	
		#p	%p	#p	%p	#p	%p	#p	%p
Sail	Sail	1	0.00%	32	0.28%	30	0.05%	103	0.17%
Kingsday	Kingsday	1	0.00%	5	0.11%	2	0.00%	5	0.01%

Druk: 'Crowded' in Dutch.
#p: Number of posts.
%p: Percentage of posts.

be derived from social media.

We also zoom-in to the two terrains, i.e. the Javakade in Sail event and Zuidplein in King's Day event. Figure 2.10 (a) and (b) shows the most frequently used words captured in these two terrains on Instagram, excluding the general words and event names. These words characterise events in these two terrains in terms of crowd sentiment and activity. For instance, frequently used words such as 'thanks', 'wonderful', and 'love' describe a very good atmosphere during the activities, mentioned in words such as 'boatlife', 'Vuurwerk' ('fireworks' in Dutch), 'sunset' in Javakade in Sail, and in music performances such as 'martingarrix' and 'TIKTAK' taking place around the Zuidplein during Kingsday.

To showcase the availability of information related to the attendees' perception which can be used for crowd managers to estimate the crowdedness of the crowd during events. We analyze the occurrence distributions of words such as 'Crowded' and 'Druk' (crowded in Dutch) in both events in Table 2.9, normalizing by the total number of posts. During both events people on Twitter posted these words more often than on Instagram: it is possible that people are more willing to share their negative emotions, such as crowded perceptions, on Twitter. While comparing two events, more 'Crowded' and 'Druk' words are captured during Sail than during King's Day. This perception is exactly what these words from social media data represent. According to Zuurbier (2019) and Duives et al. (2015), the perception of crowdedness is more important for crowd management than actual densities. This may also indicate that people experience more crowdedness during the Sail event than during King's Day.

Discussion

In the above sections, we report the outcomes of the crowd characterization operations performed on social media data that can be collected in a timely manner during the two events. These crowd properties are calculated to provide relevant insights for crowd management purposes. Crowd managers could apply crowd management measures by taking into consideration the semantic and qualitative interpretation of social media posts. Therefore, the validity and reliability of the crowd characterization is essential for crowd management.

Our work shows that, using state-of-the-art techniques, social media data can be reliably enriched to surface socio-demographic properties of their users. We acknowledge that the limited availability of meaningful profile pictures and the location information reduces the intrinsic utility of socio-demographic analysis of social media data. Although it is good to take these limitations into account, and work further on more precise methods to derive this information, the resulting information may benefit crowd management purposes, as currently hardly any information on crowd characteristics is available.

We noticed that the sparsity of social media data has less influence on characterizing the relative difference in the crowd in terms of the temporal and spatial distribution of posts, the spatial distribution of PoI popularity, and distribution of word use than on the distribution of age, gender and city-role for crowd management during city-scale events.

In contrast, the bias of social media usage affects crowd characterization for crowd management. The bias in age, e.g. old people seldom use social media, affects the identified age distribution of the crowd. The bias in gender with respect to the platform, i.e. females are more active on Instagram, affects characterizing the gender composition of the crowd. Also the bias in platform usage, i.e. the Instagram data is more sensitive to the temporal distribution than Twitter, which presents more diverse distinctions between events or between different days in an event, affects characterizing the temporal distribution of the crowd. Besides, the bias in city-role, e.g. tourists are more active than residents, reduces the capability of crowd characterization using social media. This bias can be reduced following a two-step approach. The first step is to compare the composition of social media users with population characteristics in city events from other data sources such as CBS data and counting-system data. The latter depends on the characteristics of event programmes and locations, being more specific. Therefore, the feasible way is the first step, i.e. to gain more information about the population characteristics such as sensor data, stewards observations and historical data on crowd characteristics, and reduce biases caused by social media data using bias reduction techniques (Culotta, 2014; Kuru & Pasek, 2016; Firth, 1993). In the meantime, the comparison of the distribution of age, gender, and city-role through social media in different events successfully characterized the distinction of events. E.g. the age distribution during King's Day is more uniform than during Sail. This information can be used in crowd management to gain event distinctions and apply measures accordingly.

2.5 Summary and conclusions

Nowadays, city-scale events are getting more popular. Stakeholders of these events demand qualitative and quantitative insights into the crowd to be managed. Conventional solutions depend on manual observations, which are expensive, prone to introduce observation biases, and not suitable for longitudinal observations.

In this chapter, we advocate the use of social media data as a valuable and effective alternative data source for crowd characterization purposes. We screened out a set of factors that are known to influence pedestrian behaviors which, therefore, are relevant for crowd characterization. We also select examples of methods that could be used to derive information about these factors from social media data.

We apply these methods in the context of two city scale events – Sail

2015 and King's Day 2016, in Amsterdam, the Netherlands – and reflect on the accuracy and reliability of these methods. Based on the results of the existing methods, we can conclude whether dedicated methods need to (and can) be developed. For instance, we observed that the age distribution during King's Day is more evenly distributed than during Sail, a result that complies with the expected composition of the events' crowds. We also found that less local tourists join the King's Day event in Amsterdam than during Sail, which may be explained by the fact that people in other cities are more willing to travel to Amsterdam for Sail, the event occurring once each 5-years, than the yearly King's Day. We noticed that the amount of social media posts sent by people during events is far more than during regular days, which is in line with the fact that it is more crowded during the city events. The temporal distribution of sub-events which take place in large areas, lasts for several hours during the day, e.g. the Sail-in parade and King's Day boat parade, illustrates clearly the increase of participants. Moreover, more social media usage and PoI visits are observed in IJhaven area during Sail than during the King's Day event, which is in line with the fact that Sail activities take place around the IJhaven area where ships docked. The word use of the people across events from social media successfully captured the words about event topic and people emotion. All of these observations indicate crowd management strategies could take into consideration the different characteristics of crowds, in terms of demographics, spatial-temporal distribution, Point of Interest (PoI) preferences and word use, in the context of city events.

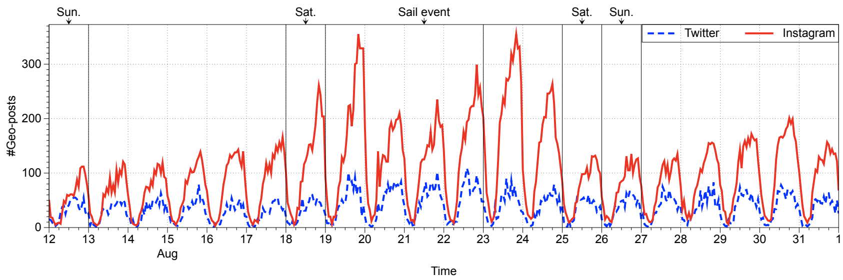
The social media sparsity has less impact on the crowd characterization. However, bias of social media usage in terms of age, gender and city-role as well as social media platform selection affects providing information about crowd during events, such as absolute age, gender distribution and temporal distribution of a crowd. Still, the comparison of results between the two events is in line with expected event characteristics.

In future work, we plan to continue studying in multiple topics based on the findings and analysis in the current study. To deal with the social media data sparsity, we would like to use geolocalisation techniques to increase amount of geo-referenced posts to counter the sparsity of geo-referenced data on social media. To provide more relevant information for crowd management, we plan to explore methods to derive more social media proxies about people in the crowd, such as their sentiment, the main area they visited in different events. A further sensitivity analysis would be done in terms of the number of points and the radius (i.e. the ”epsilon” parameter for the DBSCAN algorithm) for expanding clusters. We also would like to investigate the feasibility of the current study in other events which may have different characteristics.

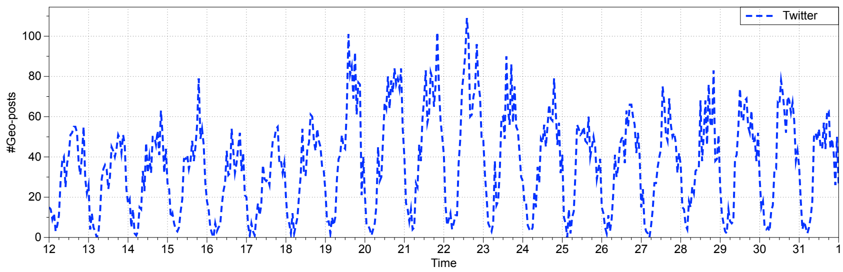
2.6 Appendix

Social media geo-posts sent around Sail and Kingsday event

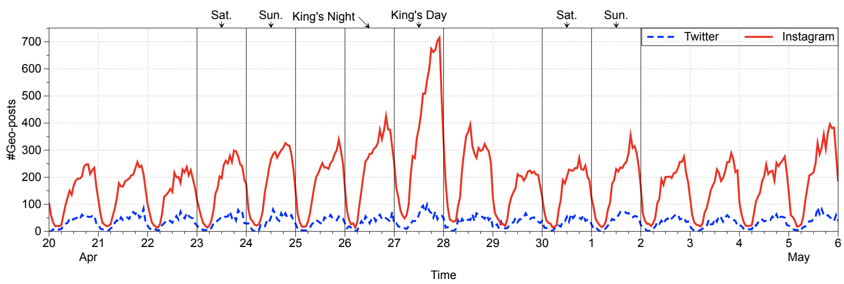
- The temporal distribution of posts sent by people observed from social media in Figure 2.11.
- The Total, Max, and Minimal number of geo-posts sent around Sail event 2015 in Table 2.10 and around King’s Day event 2016 in Table 2.11.



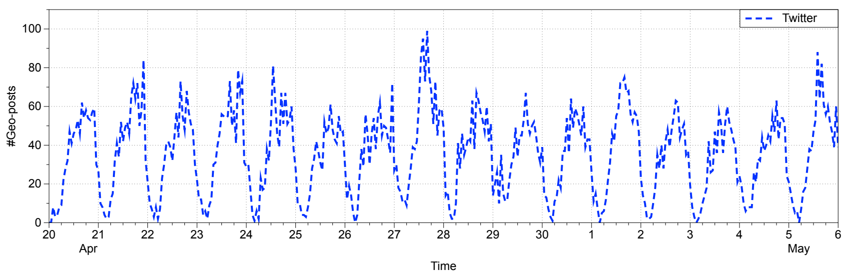
(a) Sail



(b) Sail, Twitter



(c) Kingsday



(d) Kingsday, Twitter

Figure 2.11: The temporal distribution of posts sent by people observed from social media

Table 2.10: The Total, Max, and Minimal Number of Geo-posts Sent Around Sail Event 2015

		Date																			
								Weekend	Sail Event						Weekend						
		Sun Aug 12	Mon Aug 13	Tue Aug 14	Wed Aug 15	Thu Aug 16	Fri Aug 17	Sat Aug 18	Sun Aug 19	Mon Aug 20	Tue Aug 21	Wed Aug 22	Thu Aug 23	Fri Aug 24	Sat Aug 25	Sun Aug 26	Mon Aug 27	Tue Aug 28	Wed Aug 29	Thu Aug 30	Fri Aug 31
Twitter	#P. of day	759	686	727	757	591	629	754	1083	1092	1154	1212	1002	892	834	868	832	926	905	980	836
	Max #P., Time (hour)	15:00	18:00	20:00	19:00	15:00	19:00	15:00	14:00	18:00	20:00	14:00	14:00	19:00	16:00	17:00	13:00	20:00	12:00	13:00	17:00
	Max #P.	55	51	63	79	54	55	61	101	84	102	109	90	79	60	66	75	83	74	78	63
	Min #P., Time (hour)	04:00	05:00	05:00	06:00	04:00	05:00	03:00	03:00	04:00	05:00	05:00	05:00	05:00	03:00	03:00	05:00	04:00	04:00	06:00	05:00
	Min #P.	3	0	1	3	0	1	2	2	1	4	4	3	3	0	2	0	4	2	3	2
Instagram	#P. of day	1291	1435	1510	1696	1943	2012	2796	3775	2967	2766	3429	4638	3278	1796	1751	1815	2156	2402	2754	2268
	Max #P., Time (hour)	19:00	21:00	21:00	20:00	19:00	22:00	20:00	20:00	21:00	20:00	20:00	20:00	18:00	20:00	18:00	22:00	19:00	19:00	18:00	17:00
	Max #P.	112	121	119	139	143	166	261	355	210	235	299	355	264	132	135	135	156	172	201	157
	Min #P., Time (hour)	03:00	05:00	05:00	05:00	05:00	04:00	05:00	04:00	04:00	06:00	06:00	05:00	05:00	04:00	06:00	05:00	06:00	05:00	06:00	05:00
	Min #P., Time (hour)	5	4	4	5	3	5	5	6	12	9	8	11	16	9	10	10	16	6	12	9

#GP.: amount of Geo-posts.
Max #GP. Time(hour): the time in hour during which the max amount of Geo-posts is observed.

Table 2.11: The Total, Max, and Minimal Number of Geo-posts Sent Around King's Day Event 2016

		Date															
					Weekend			K. Night	K. Day			Weekend					
		Wed April 20	Thu April 21	Fri April 22	Sat April 23	Sun April 24	Mon April 25	Tue April 26	Wed April 27	Thu April 28	Fri April 29	Sat April 30	Sun May 1	Mon May 2	Tue May 3	Wed May 4	Thu May 5
Twitter	#P. of day	822	926	839	918	886	768	855	1134	914	833	814	957	785	715	794	909
	Max #P., Time (hour)	16:00	22:00	16:00	20:00	13:00	17:00	23:00	16:00	16:00	16:00	14:00	16:00	17:00	18:00	18:00	14:00
	Max #P.	62	84	73	79	81	61	72	99	67	67	64	75	63	60	63	88
	Min #P., Time (hour)	01:00	04:00	05:00	05:00	04:00	05:00	05:00	06:00	04:00	06:00	05:00	04:00	04:00	03:00	04:00	05:00
	Min #P.	0	2	2	2	1	3	0	9	2	10	1	0	2	0	6	0
Instagram	#P. of day	3369	3402	3325	4006	4574	4277	5450	8902	5575	3407	3622	4160	3666	3647	3723	4959
	Max #P., Time (hour)	19:00	19:00	19:00	18:00	19:00	21:00	20:00	22:00	13:00	18:00	18:00	19:00	21:00	20:00	22:00	20:00
	Max #P.	248	256	238	299	326	339	425	715	394	225	271	355	276	289	276	396
	Min #P., Time (hour)	03:00	04:00	05:00	05:00	05:00	04:00	05:00	05:00	05:00	06:00	05:00	05:00	05:00	05:00	05:00	04:00
	Min #P.	19	19	14	14	22	17	14	48	36	20	15	16	20	18	20	18

Areas of Interest based on clustering social media posts

The Area of Interest detected by clustering PoIs using social media posts for 4 sub-events listed in Table 2.8 are shown in Figure 2.12, 2.13, 2.14, and 2.15. These cluster maps are generated using DBSCAN algorithm (Ester et al., 1996), which groups points within distance of 50 meters (epsilon=0.05). Such groups will be identified as a cluster if the number of points in a group is more than 15 (min_Points = 15), otherwise each points will be determined as outliers. Each colour represents one cluster. Black points are outliers which fail to form group with any other points. The count number denotes the amount of points in each cluster.

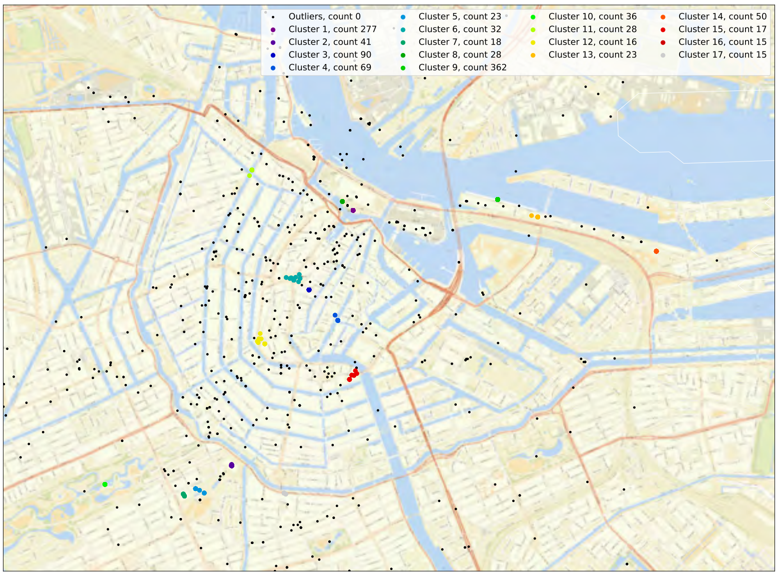


Figure 2.12: The Area of Interest detected by clustering PoIs using social media posts in Sail, Sail-in Parade

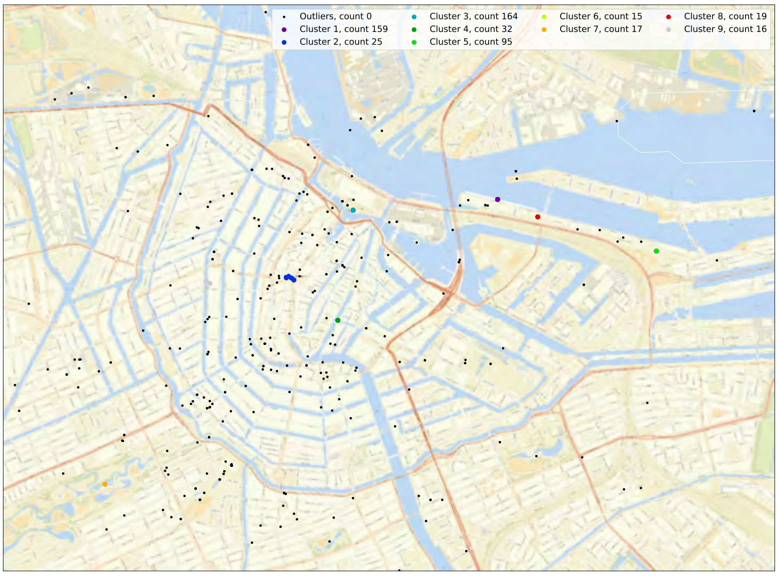


Figure 2.13: The Area of Interest detected by clustering PoIs using social media posts in Sail, Fireworks

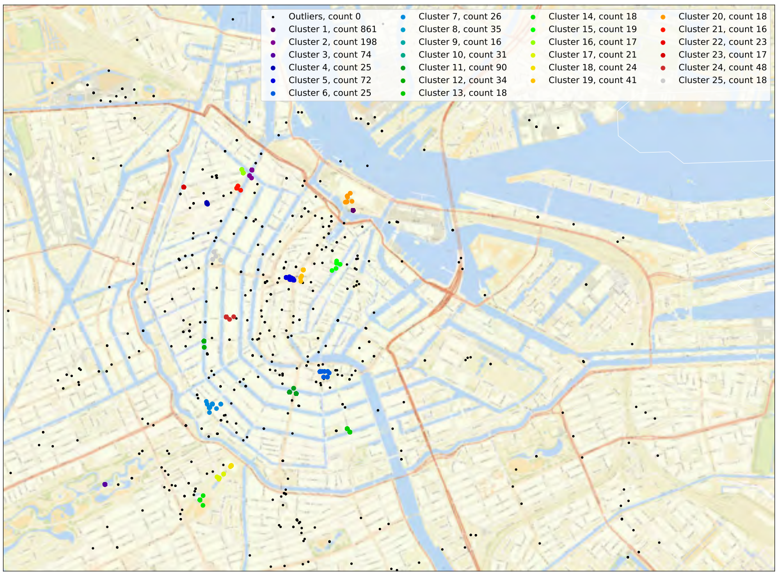


Figure 2.14: The Area of Interest detected by clustering PoIs using social media posts in King's Day, Boat Parade

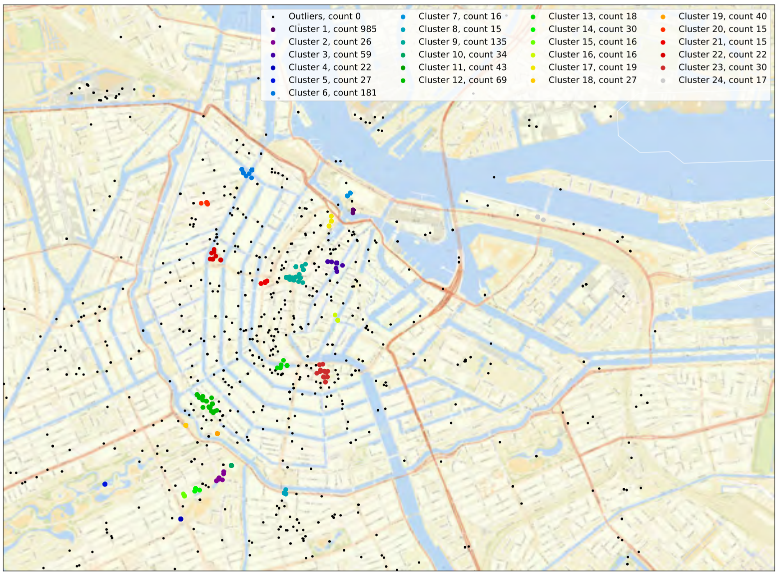


Figure 2.15: The Area of Interest detected by clustering PoIs using social media posts in King's Day, King's Night

Chapter 3

Estimate Sentiment of Crowds from Social Media during City Events

In the previous chapter, we explored crowds characterisation using social media data in terms of different aspects such as crowd demographics, city-role, temporal distribution, post position, Points of Interests, and word use. In this chapter, we aim to characterise the attendees' emotions during the events. We derive and analyse the sentiment of people using social media text. This answers the second research question, i.e. **RQ2. To what extent are social media data able to estimate the sentiment of crowds in city events?**

We first selected some lexicon and machine learning methods to perform sentiment analysis. Next, we constructed an event-based sentiment annotated dataset. We trained and tested the performance of the selected methods to estimate the sentiment from text using common and event-based datasets.

Results show that machine learning based (ML-based) methods show better performance than lexicon-based methods in most situations in terms of sentiment estimation. In particular, the ML-based method Linear Support Vector Classifier (LinearSVC) reaches the minimal estimation error at approximately 0.177 when trained and tested with an event-based dataset constructed using social media data collected during city events. The proposed event-based dataset is essential for training methods to reduce estimation error in such contexts.

This chapter is published as a journal article: Gong, V. X., Daamen, W., Bozzon, A., & Hoogendoorn, S. P. (2019). Estimate sentiment of crowds from social media during city events. *Transportation research record*, 2673(11), 836-850.

3.1 Introduction

As cities compete for global attractiveness and community quality, city-scale public events become more and more popular to boost tourism and promote economic growth.

Thematic exhibitions, sports competitions, and national celebrations are instances of city events that take place in urban areas, and may attract a large amount of people during a short time period. The scale and intensity of these events request technical solutions supporting stakeholders (e.g., event organisers, public and safety authorities) to manage the crowd.

The crowd during an event is managed by public authorities to reduce risks of incidents as results of internal and external threats. This is usually achieved by exerting predefined measures based on qualitative interpretations of the crowds by stewards, police officers, or event organization employees.

As the efficiency and effectiveness of crowd management measures depend on pedestrian behavior (Still, 2000; Zomer et al., 2015), it is beneficial for stakeholders to obtain information about the behavior of the crowd. The sentiment of people in the crowd is one of the factors affecting crowd behavior (Martin, 2006). Together with other information such as crowds density and demographics, it may help crowd managers estimate and predict (negative) behaviours that can be inferred from the sentiment of people in the crowd, such as risky behaviours. Therefore, deriving the sentiment of people in the crowd could be valuable to crowd management.

However, the sentiment of crowds is difficult to acquire. Conventional approaches capture this information manually by stewards or staff members (Earl et al., 2004), a practice that is costly and subject to bias. Traditional crowd observation techniques are based on sensors (e.g., counting systems, GPS trackers and Wi-Fi sensors) which only provide spatio-temporal information. These solutions do not provide sentiment values. Although crowd sentiment could be extracted from image or video clips provided by cameras through image recognition techniques (Poria et al., 2016b,a), accessing the images or video recordings of public area is computationally intensive, and often restricted due to privacy issues.

The advance of Web-based technologies provide new data sources that could be applied to understand and analyse pedestrian behavior (Cranshaw et al., 2012; Quercia et al., 2015a; Hasan et al., 2013; Gong et al., 2018b). Several social media networks, such as Twitter and Instagram, are widely used. Time-stamped social media posts, such as text content, are often geotagged. More importantly, these posts intrinsically embrace rich semantic information that could be employed for deriving crowd sentiments. Therefore, social media data can be used to derive the sentiment of crowds during events.

A large number of works studied sentiment estimation of crowds using social media data. Jiang et al. (2011) introduced a method to estimate sentiment of tweets considering multiple strategies and including context information (i.e., related tweets). Zhang et al. (2014) demonstrated a machine-learning method incorporate syntactic and context information from social media to estimate users sentiment. Ortigosa et al. (2014) presented a method to estimate sentiment of users based on their Facebook texts, which integrates lexicon-based, machine-learning techniques.

Yu et al. (2013) applied the sentiment analysis to financial sector, where they derive sentiment of traders from social networks, and explore its relationship with short term stock market performance. Surveys about sentiment analysis (Pang et al., 2002; Ravi & Ravi, 2015) reviewed more than 20 methods and 30 works estimating sentiment from social media. Those methods can be categorised into three types: lexicon based methods (lexicon-based), machine learning based methods (ML-based methods) and hybrid methods. Lexicon-based methods assign each consecutive combination of words of a text a sentiment score according to a dictionary and calculate the weighted average sentiment score. The ML-based methods train the model with a sentiment annotated dataset and estimate the sentiment of a test dataset through the model. Hybrid methods are a combination of lexicon- and ML-based methods. With respect to the context, sentiment analysis in context of city events for crowd management differ from other contexts (e.g. E-learning, marketing, stock prediction) in a set of characteristics, such as the specific topic of the event, its location, popularity, and time of occurrence. Consequently, sentiment analysis methods suitable for other contexts may differ with methods fit for context of city events. While showing the utility of social media data in sentiment analysis studies, no previous work aimed at deriving sentiment of crowds in the context of city events. Moreover, the sentiment annotated dataset for a specific context is significant in sentiment analysis. It can be used for evaluating sentiment estimation result from various methods. It can also be used for ML-based or hybrid methods to train their model for sentiment estimation. However, none of previous works proposed sentiment annotated datasets in such a context. What is missing is an in-depth understanding of which methods are most effective in this context, and whether their performance will be affected by the diversity of the events or the urban areas they take place.

These research gaps lead to the following research question: Which methods are suitable to derive sentiments of crowds from social media texts in city events?

To answer this research question, we selected a number of methods. In order to compare and assess their performance in the context of city events, an event-based sentiment annotated dataset was required as ground truth. As no annotated event-based datasets exist, we constructed one. Using this dataset, we tested the performance of candidate methods, and selected the most promising method.

We present a literature review in the next section, followed by the research methodology to examine the performance of candidate methods. Then, the methods for comparison are selected, followed by a description of the data collection. Further, we introduce the experiment setting, followed by the findings, analysis and discussion about the experiment result. The conclusion and future work is included in the end of the paper.

3.2 Literature Review

In our work, we compare sentiment analysis performance of various methods and propose an event-based sentiment dataset. In this section, we briefly review previous works about comparison of sentiment analysis methods and the proposed sentiment datasets.

Previous works perform experiments in certain context or for certain purposes, such as for document classification, e-learning and brand marketing. They select a set of methods for comparison, use datasets to train their model and evaluate the results. Therefore, three elements involved, i.e. the context for comparing methods, the selected methods for comparison, and the datasets for training and testing. In the following, we look into previous works with regard to above three elements.

Sentiment analysis has been applied in various context. Pang et al. (2002) in their work investigate a set of methods to estimate sentiment for classifying documents. They compare the sentiment analysis performance of several machine learning methods based on public reviews collected from Internet, such as movie reviews from IMDB[1]. Selected methods for comparison include Naive Bayes (NB), Support Vector Machines (SVM) and Maximum Entropy (ME). Result shows that SVM outperforms other methods in their experiment. Boiy & Moens (2009) apply sentiment analysis for opinion mining on multilingual web texts. They analyse the performance of machine learning methods including NB, SVM and ME to estimate sentiment in public reviews about cars and movies. Findings show that SVM outperforms other methods. Li & Li (2013) estimate sentiment on social media for marketing purpose. They derive market intelligence through sentiment analysis based on product review on Twitter, such as reviews for Microsoft, Sony, iPhone, iPad and Macbook. They compare performance of NB and SVM with various settings. Findings show that SVM reaches the better performance in their experiment. Ortigosa et al. (2014) analyse sentiment in the context of e-learning based on social media data. A set of machine learning methods including a hybrid method are compared for sentiment analysis on text from Facebook. In contrast to other works, they find that NB outperforms other methods including SVM in this context. Bravo-Marquez et al. (2014) study the sentiment analysis on big social data. They compare sentiment through various methods with different configuration based on a large volume non-topic specific tweets. They also find that NB shows better performance than SVM in their experiment.

Above works show that NB and SVM are the most popular machine learning methods for sentiment analysis (Pang et al., 2002; Boiy & Moens, 2009; Li & Li, 2013; Thelwall et al., 2010, 2012; Bravo-Marquez et al., 2014; Ortigosa et al., 2014). Public reviews and social media data are widely used for deriving sentiment. Some investigations (Pang et al., 2002; Boiy & Moens, 2009; Li & Li, 2013; Ortigosa et al., 2014) are performed with data related to certain topics, such as reviews of movies, cars, as well as brands, e.g. Microsoft, Sony and Google. While, other works (Thelwall et al., 2010, 2012; Bravo-Marquez et al., 2014) are performed with common-based dataset. However, none of these works is performed in context of city event with respect to crowd management, and none of these works investigate the impact of using different datasets, i.e. common-based or a certain topic based, on the performance of sentiment analysis.

Analogues to datasets construction, Kouloumpis et al. (2011) publish

[1] IMDb (Internet Movie Database) is an online database of information related to films, television programs, home videos and video games, and internet streams and fan reviews and ratings. https://www.imdb.com/

a dataset including 222,570 sentences annotated with three sentiment categories, i.e. positive, neutral and negative. These sentences are collected from Twitter with no specific topic. Costa et al. (2014) publish an comprehensive datasets containing 400 deceptive and 400 truthful reviews in positive and negative category. These datasets are widely used in several sentiment analysis works (Liu, 2012; Kiritchenko et al., 2014; Saif et al., 2012; Zhang & Chow, 2015; Aydoğan & Akcayol, 2016). Besides, there are datasets proposed related to a certain topic. For instance, Hu & Liu (2004) present a datasets with 6,800 opinion words on 10 different products. Cruz et al. (2013) propose a sentiment annotated reviews with different topics, i.e. 587 reviews on headphones, 988 reviews on hotels and 972 reviews on cars. Similarly, Blitzer et al. (2007) propose a sentiment annotated dataset including Amazon reviews on 4 domain, i.e. books, DVDs, electronics, kitchen appliances. However, there is no dataset proposed for sentiment analysis in context of city events.

3.3 Research Approach

In this section, we elaborate the research approach for testing the sentiment estimation performance of candidate methods in the context of city events using social media.

The research approach is illustrated in Figure 3.1. It consists of three major steps, i.e. estimate the sentiment using different methods and different datasets, evaluate the estimation error per method, and compare the estimation error between methods. It involves two different datasets, i.e. the common and event-based dataset. Common-based datasets in this research cover a wide variety of situations and have no domain knowledge or context information about city events. Moreover, these datasets are well-known in the research area of sentiment analysis. While event-based datasets refer to social media posts collected during events with sentiment annotated. The event-based dataset is generated during this research. In the first step, lexicon-based methods estimate sentiment of each text in common and event-based dataset, respectively. The ML-based methods train their model using part of the common and event-based dataset. Then both trained models estimate sentiment of each text in the rest of common and event-based datasets, respectively. In this step, both methods yield estimated sentiment. In the second step, to verify whether the estimated sentiment is correct, the estimation result is compared to the sentiment ground truth of the test dataset using the metrics introduced in the next paragraph. We perform an N-round testing to reduce the random error. In each round, we select a subset, identify sentiment of each text and compare those estimated sentiment with the ground truth. Finally, we compare the performance of methods across different datasets. The analysis results also provide feedback on the research methodology, such as to adjust the sample size of the training and testing dataset, to select feasible candidate methods, and to choose suitable comparison metrics.

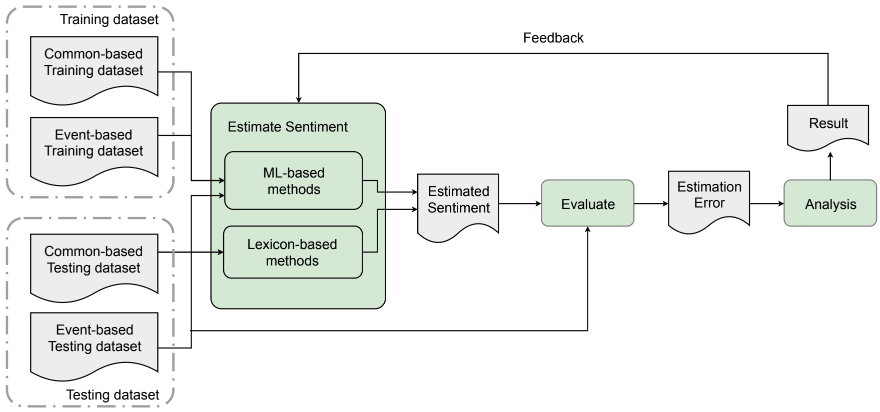


Figure 3.1: The process of investigating sentiment estimation performance of selected methods on social media text in city events. The green symbols denote 3 major steps, i.e. estimate the sentiment using different methods with different datasets, evaluate the estimation error per method, and analyse the estimation error between methods. The gray symbols denote documents input and output.

3.3.1 Comparison metrics

The sentiment estimation performance is assessed using the estimation error, which is calculated for each repetition of sentiment estimation. This estimation error per repetition E_i is calculated by the amount of false identifications M_i^{false} divided by the testing sample size M_i, see Eq. 3.1. While running an N-round testing, the mean and standard deviation of the estimation error is calculated in each round.

$$E_i = \frac{M_i^{false}}{M_i} \tag{3.1}$$

3.4 Selection of Candidate Methods from Literature

As indicated in the introduction section, the sentiment of the crowd is one of the factors affecting crowd behaviour (Martin, 2006), which can be used by crowd managers to estimate and predict (negative) behaviours of the crowd such as risky behaviours during city events. To derive the sentiments of the crowd, we compare in this section the effectiveness of various sentiment analysis methods. As indicated in the introduction section, there is no existing literature comparing the performance of sentiment analysis methods using social media in the context of large-scale city events. Thus, we review sentiment analyses methods applied for generic situations.

Deriving sentiments from social media text is not a novel problem. Many papers discussed this topic and proposed methods to solve this problem (Pang et al., 2008; Martínez Cámara et al., 2011; Ravi & Ravi, 2015). 161 studies have been reviewed in Ravi & Ravi (2015); about 30 papers discuss sentiment analysis on social media networks. As introduced in previous section, the sentiment analysis methods can be categorised into three types: lexicon based methods (lexicon-based), machine learning based methods (ML-based) and hybrid methods. Hybrid methods are a combination of lexicon- and ML-based methods. The performance of hybrid methods is therefore influenced by the quality of the lexicon- and ML-based methods it consists of. Understanding the performance of lexicon- and ML-based methods is necessary to investigate hybrid methods. Therefore, we focus in our research on lexicon- and ML-based methods, and leave hybrid methods for future work.

The overview of methods selected for sentiment estimation on social media text in the context of city events is shown in Table 3.1. More details on the selected methods are introduced in the following.

3.4.1 Lexicon based methods

Lexicon-based, or dictionary based, approaches are widely applied in the field of sentiment analysis (Ravi & Ravi, 2015; Montoyo et al., 2012). Given

Table 3.1: Selected methods for deriving sentiment of crowd on social media in city events

Category	Method		Description	Linear or Nonlinear
Lexicon-based	SentiStrength		Optimised for social media text	n/a
	SentiWordNet		Assigns to WordNet synset	n/a
ML-based	Naive Bayes (NB)		Bayes theorem	Linear
	SVM	SGDClassifier	Fitted with Stochastic Gradient Descent learning	Linear
		LinearSVC	Linear Support Vector Classification	Linear
		NuSVC	Statistical, Nu-Support Vector Classification	Nonlinear
		SVC	Statistical, C-Support Vector Classification	Nonlinear

a text from a social media post, lexicon-based methods assign each n-gram (i.e. consecutive combination of words) a sentiment score according to its attached dictionary and calculate the weighted average sentiment score as a performance indicator after filtering out stop words and reducing other noises. More than 41 studies explore lexicon-based methods in sentiment analysis, according to Ravi & Ravi (2015). Among those, SentiStrength and SentiWordNet are two popular lexicon based methods used for deriving sentiment from social media data.

SentiStrength

SentiStrength was created by identifying sentiments expressed in the texts on MySpace, a social media platform. It estimates the strength of negative, neutral and positive sentiment in short texts, originally developed for the English language and optimized for short social media texts (Thelwall et al., 2010). SentiStrength reports three sentiment strengths: -5 to -1 as negative, 0 as neutral, and 1 to 5 as positive. It has been applied and investigated in many papers in which it shows significant performance (Thelwall, 2013; Thelwall et al., 2011; Pfitzner et al., 2012; Thelwall et al., 2012).

SentiWordNet

SentiWordNet is a lexical resource for opinion mining. Instead of constructing its sentiment dictionary from a corpus (e.g. MySpace data) as the SentiStrength does, it assigns to each syncset of WordNet three sentiment scores: positive, negative, objective (Baccianella et al., 2010). It is widely used in estimating sentiment from social media networks (Nakov et al., 2016; Gilbert, 2014; Kiritchenko et al., 2014; Thelwall et al., 2010).

3.4.2 Machine Learning based methods

Machine Learning methods (ML-based methods) train the model with a sentiment annotated dataset and estimate the sentiment of a test dataset through the model. A large amount of ML-based methods is proposed and investigated in recent studies in various situations (Ravi & Ravi, 2015). Among these methods, the Naive Bayes (NB) and Support Vector Machines (SVMs) are widely tested and outperform most of other methods while deriving sentiments from social media texts (Pang et al., 2002; Ravi & Ravi, 2015; Martínez Cámara et al., 2011). In the following these methods are described in more detail.

Naive Bayes (NB)

Naive Bayes is a supervised linear machine learning algorithm which is popular to classify text. It is a simple probabilistic classifier based on applying Bayes' theorem (Murphy et al., 2006). It is widely used to estimate sentiments from social media texts (Ortigosa et al., 2014; Rui et al., 2013; Bravo-Marquez et al., 2014). Although its mechanism is fairly straightforward, it often performs as good as much more complicated solutions (Abdul-Mageed et al., 2014; Yu et al., 2013).

Support Vector Machines (SVMs)

Support Vector Machines are a family of supervised learning models used for linear and non-linear classification analysis. SVMs are widely used in text categorisation for sentiment analysis (Pang et al., 2002; Ravi & Ravi, 2015). In this research, we test the four most popular SVM models, namely Stochastic Gradient Descent Classifier fitted SVM (SGDClassifier, or SGDC), Linear Support Vector Classifier (Linear-SVC, LinearSVC or LSVC), Nu-Support Vector Classifier (NuSVC) and Support Vector Classifier (SVC). The SGDClassifier is a linear SVM classifier fitted with stochastic gradient descent (SGD) learning. LinearSVC is an implementation of Support Vector Classification (SVC) in case of a linear kernel. SVC and NuSVC apply the statistics of support vectors, developed in the support vector machines algorithm. SVC and NuSVC are similar methods, but accept slightly different sets of parameters and have different mathematical formulations. These methods are explored in a large number of papers deriving sentiment from social media (Tripathy et al., 2016; Nabil et al., 2015; Al-Azani & El-Alfy, 2017; Jiang et al., 2011; Maas et al., 2011; Rui et al., 2013; Ortigosa et al., 2014; Abdul-Mageed et al., 2014; Thelwall et al., 2010; Bravo-Marquez et al., 2014).

3.4.3 Sentiment estimation result scheme

In this research, we are aiming at comparing the performance of sentiment analysis methods. However various selected methods result in different sentiment schemes. For instance, the lexicon-based method SentiStrength outputs sentiment values in an integer between -5 and 5, while the SentiWordNet results in negative, neutral and positive. For other methods, the output schemes are listed in Table 3.2. To compare the performance of these selected methods, it is necessary to define a unified output scheme and map all schemes of those selected methods to the unified output scheme.

According to Table 3.2, there are two types of sentiment scheme, i.e. the simplified one and the detailed one. The simplified scheme in this research refers to the sentiment scheme featuring only 3 categories, i.e. positive (1), neutral (0), and negative (-1). Whereas, a detailed sentiment scheme may have more sentiment categories, e.g. extremely negative, very negative, negative, slightly negative, neutral, slightly positive, positive, very positive, extremely positive.

For lexicon-based methods, SentiStrength supports a detailed sentiment scheme, while SentiWordNet results in the simplified scheme. For ML-based methods, the supported sentiment scheme depends on the training data. Namely, if the training dataset is annotated with a detailed sentiment scheme, the ML-based methods trained with such dataset also yield sentiment score in the same scheme.

However, when constructing such dataset, the agreement reached on a sentiment category from a detailed scheme, e.g. extremely negative, is less than on a category from the simplified scheme, e.g. negative. Moreover, subjective errors introduced by the human being in the annotation is also increased when using the detailed scheme. Thus, a dataset annotated with a detail sentiment scheme is difficult to construct, less reliable, and therefore more rare. Most of the existing sentiment datasets are annotated with a simplified scheme, i.e. negative, neutral and positive. ML-based methods trained with such dataset also result in such simplified sentiment scheme.

With regard to the impact of simplified sentiment scheme on the estimation error of the models, compared with a detailed sentiment scheme, a simplified one indeed may lose the detailed sentiment strength information, but still it reports the same sentiment polarity, e.g. either very positive or slightly positive in a detailed scheme will be reported as positive in a simplified scheme.

As the simplified sentiment scheme is widely supported by both lexicon and ML-based methods, we apply the simplified sentiment score in this research. We assign the following three sentiment values: -1, 0 and 1, denoting negative, neutral and positive respectively. The mapping of all selected methods is shown in Table 3.2.

Table 3.2: Sentiment estimation output schema transformation rules

Category	Name	Output schema	Convert rule	Unified Schema
Lexicon-based	SentiStrength	-5 to -1 as negative, 0 as neutral, 1 to 5 as positive	-5, -1] as -1, 0 as 0, [1, 5] as 1	-1 denotes Negative, 0 denotes Neutral, 1 denotes Positive
	SentiWordNet	negative, neutral, positive	negative as -1, neutral as 0, positive as 1	
ML-based	NB SGDClassifier LinearSVC NuSVC SVC	-1, 0, 1 or -1, 1	-	

3.5 Data Collection

Investigating the performance of candidate methods in the context of city events requires ground truth data, both for testing purposes, and, in case of ML-based methods, to train their model. In our research, we need both common-based and event-based sentiment annotated social media data. Annotation in this respect means that for each text, its sentiment is known. Common-based datasets cover a wide variety of situations and have no domain knowledge or context information about city events, while event-based datasets focus on posts have been collected during events. As indicated in the introduction, annotated common-based datasets are available from previous research, but annotated event-based datasets are not yet available. To fill this research gap, we construct such an event-based dataset.

The common and event-based sentiment dataset is annotated with sentiment polarities, i.e. Positive, Neutral and Negative. As activities in city events, e.g. celebrations and riots, tend to stimulate attendees' sentiment. The sentiment of crowds in the context of city events are stronger than in a normal context, namely more extreme (positive or negative) expressions than neutral ones. Thus, it is valuable to explore the distinction of sentiment estimation with and without neutral polarity. In this research, we consider sentiment polarity in two sets; one consists of Positive and Negative denoted as PN, and another with Positive, Neutral and Negative denoted as PNN. The sentiment analysis with and without other individual polarities will be kept for future research.

In this section, we describe the selection of the common-based datasets and the construction of an annotated event-based dataset.

3.5.1 Common-based dataset

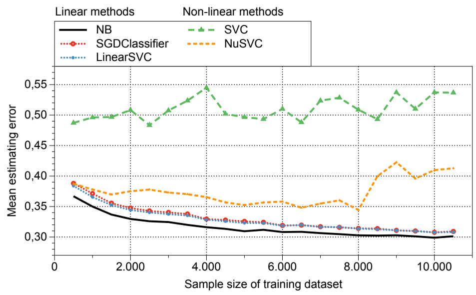
There is no official definition for the common-based dataset. In this research, it refers to datasets that cover a wide variety of situations, and contains no domain knowledge or context information about city events. Several papers proposed sentiment annotated datasets consisting of texts collected on social media. The most comprehensive paper (Ravi & Ravi, 2015) listed 32 public datasets used for sentiment analysis, 6 of which are social media datasets.

These datasets vary in terms of topic, sentiment polarity and annotation approaches. Social media posts contained in these datasets may cover diverse topics, such as digital brands, sports and technology. With regard to sentiment polarity, some datasets contain posts with Positive, Neutral and Negative polarities, while others only have Positive and Negative posts. There are two major annotation approaches, i.e., by authors themselves or through crowd sourcing. As our common-based ground truth, we choose two social media datasets, given their large amount of texts, the fact that they are widely used in other research, their diverse sentiment polarities and topics. The first one is the dataset for the University of Michigan Sentiment Analysis competition, which consists of more than 1.5 million social media posts annotated with Positive and Negative. It is applied several times as the ground truth for this competition. The second dataset is an extended version of Niek Sanders sentiment dataset series which are widely explored in sentiment analysis studies (Pandey et al., 2017; Lima et al., 2015; Aston et al., 2014; Gurkhe et al., 2014). This dataset contains more than 55,000 social media posts, each of which annotated with Positive, Neutral or Negative. Posts in both datasets cover random topics.

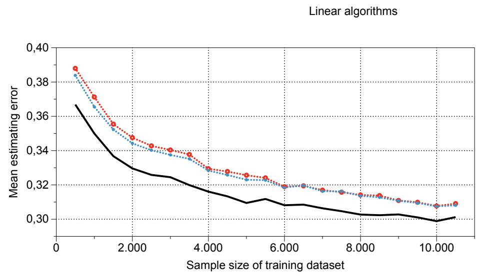
3.5.2 Event-based dataset

Event-based datasets in this research refer to sentiment annotated datasets consisting of social media posts posted during both city-scale events and local events. It should be sufficiently large to be used as ground truth for testing candidate methods, but also serve as training data for ML-based methods. To construct such an event-based dataset, we first estimated the required size of the event-based dataset. Then, we collected social media posts and annotated them with sentiment scores. This is elaborated upon in the following.

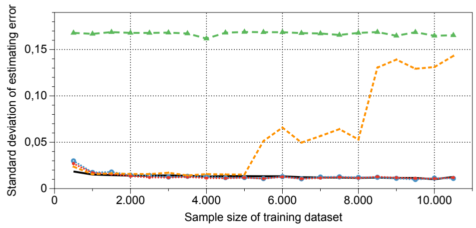
The size of the event-based dataset should meet two criteria. First, it should be sufficient as training dataset for ML-based candidate methods to reach stabilised performance for sentiment analysis. Second, it should be as small as possible given the efforts and costs to perform the sentiment annotation. To estimate the sample size, we used the common-based dataset to investigate the estimation performance variance for different sample sizes using different ML-based methods, shown in Figure 3.2. Figure 3.2(a) shows the variance of mean estimation error with respect to the size of the training sample. As expected, the mean error of Linear ML-based methods decreases when the training sample is increased. However, the nonlinear ML-based methods shows unexpected increases with increasing sample size, which may be caused by their nonlinear nature. In order to present the variation pattern of the linear method more clearly, we zoom in on the error in Figure 3.2(b). The figure shows that the mean estimation errors of linear ML-methods decrease considerably when the training sample is less than 2000 posts, after which the decrease in error becomes less sharp and gradually flattens after a sample size of 6000 posts. Taking a training sample size of 6000, increasing the sample with another 2000 posts only reduces the mean estimation error with less than 0.01. This also holds for the decrease of the standard deviation of estimation error, shown in Figure 3.2(c) and 3.2(d). We can thus conclude that the estimation error of linear ML-based methods stabilises when the training sample is higher than 6000, while training sample with more than 6000 posts do not reduce the estimation error significantly. Thus, we choose 6000 as the training sample size. In addition, we need around 15% of the dataset to test the method performance, so the event-based dataset should contain around 8000 posts.



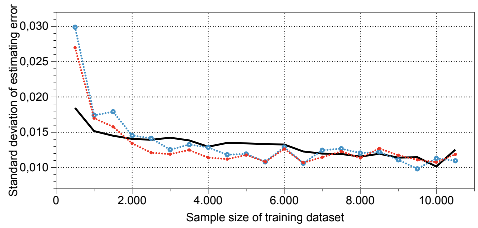
(a) Mean Estimating Error



(b) Mean Estimating Error - Linear Methods



(c) Standard Deviation Estimating Error



(d) Standard Deviation Estimating Error - Linear Methods

Figure 3.2: The mean and standard deviation of estimation error when increasing the training data sample size for ML-based methods. (b) and (d) zoom in linear ML-based methods to present the variation pattern of linear methods more clearly. The estimation error of linear ML-based methods stabilises when the training sample size is higher than 6000.

We construct the event-based dataset following the process shown in Figure 3.3. After estimating the size of the dataset, we identified the requirements regarding the events and activities considering diversity in cities, event characteristics, and their major activities.

To select the cities and activities we identified several criteria, listed below.

- Different cities. We decided to use social media posts of the two big cities in the Netherlands, i.e., Amsterdam and Rotterdam, as they will provide us with sufficient posts and cover the slightly different population as well.

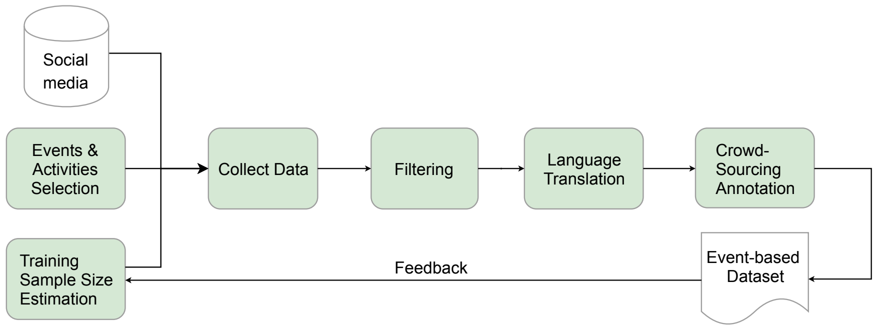


Figure 3.3: The process of constructing the annotated event-based social media sentiment dataset.

- Different event characteristics. We selected four events, listed in Table 4.3.
  - A large nautical event (Sail 2015)
  - An annual national holiday (Kingsday 2016)
  - An annual celebration includes a canal parade and parties (Europride 2017)
  - Football riots and championship celebration (Feyenoord 2017)
- Different activities during these events, including:
  - Canal parade
  - Street parties
  - Flea market
  - Fireworks
  - Riots
- Different areas in the city. Here, we vary between events that spread out over the whole inner city and events that are located in a smaller area.

Then we collect geo-referenced tweets and Instagram posts from selected city events according to the estimated sample size. The collection of social media posts during these events is performed through the API of social media platforms with the help of SocialGlass [2], an integrated system for collecting and processing social media data. Further, we filter out spam accounts and short posts (i.e., length smaller than 30 characters) which may contain useless or insufficient information for sentiment analysis. As city events attract many foreigners, posts may contain various languages, rather than only English. To determine the sentiment of those posts, we translated all posts into English using the Google Translate API [3] as it provides acceptable results compared with other translation services (Aiken & Balan, 2011). The sentiment of each post is then annotated through crowd-sourcing: the sentiment of each post is determined by multiple people and the majority judgement is taken as the ground truth. We perform the crowd-sourcing operation using Figure Eight[4], a popular crowd-sourcing platform. Each post is annotated using one of the terms positive, neutral or negative.

The characteristics of the common and event-based datasets used in this study are presented in Table 3.4. For each category we have two datasets, i.e., one with sentiment polarity of Positive and Negative, the other with the polarities Positive, Neutral and Negative.

Table 3.3: City events and activities for constructing event-based dataset

Event	Activity	Date			Place	
		Day	Start time	End time	Area	City
Sail 2015	SAIL-In parade	2015-8-19	10:00:00	17:00:00	IJ	Amsterdam
	SAIL Thank You parade	2015-8-23	16:30:00	21:00:00	IJ	Amsterdam
	Fireworks	2015-8-19 to 23	22:30:00	22:45:00	IJ	Amsterdam
Kingsday 2016	Kid's flea market	2016-4-27	9:00:00	18:00:00	Vondelpark	Amsterdam
	Canal boat party	2016-4-27	14:00:00	17:00:00	City center	Amsterdam
	King's Night	2016-4-26 to 27	20:00:00	2:00:00	Rembrandtplein and Melkweg	Amsterdam
Europride 2017	Canal parade	2017-8-5	13:00:00	17:00:00	City center	Amsterdam
	Pride park	2017-7-28	14:00:00	23:00:00	Vondelpark	Amsterdam
	Streetparties (Street Party) 1	2017-8-4	19:00:00	2:00:00	Reguliersdwarsstraat	Amsterdam
	Streetparties (Street Party) 2	2017-8-5	16:00:00	2:00:00		
Feyenoord 2017	Lose the match	2017-5-7	00:00:00	23:59:59	Feyenoord De Kuip stadium	Rotterdam
	Win the league title	2017-5-14	00:00:00	23:59:59	Feyenoord De Kuip stadium	Rotterdam

[2] http://social-glass.tudelft.nl/
[3] https://cloud.google.com/translate/
[4] https://www.figure-eight.com/

Table 3.4: Common and Event-based dataset for sentiment estimation

Category	Name	Sentiment Polarity	Source	# Positive	# Neutral	# Negative	# Total
Common-based	CA	Positive, Negative	University of Michigan Sentiment Analysis competition	789914	-	788127	1578041
	CB	Positive, Neutral, Negative	Niek Sanders sentiment dataset (extended)	16146	14004	25379	55537
Event-based	EA	Positive, Negative	Constructed	5040	-	2029	7069
	EB	Positive, Neutral, Negative	Constructed	5040	1093	2029	8162

3.6 Experimental Setting

In this section, we describe the set up of the experiment to test the sentiment estimation of crowds in city events using the selected methods using social media. The experiment involves multiple control variables. In the following, we first describe the values of each control variable, then introduce the experimental scenarios which combine the variable values. In the last subsection, we describe experiment setting applied in this experiment, i.e., the training and testing sample size, the number of rounds for N-round testing.

3.6.1 Control variable

The experiment is designed to test the sentiment estimation performance of selected methods in city events. It therefore consists of methods and testing datasets as variables. Besides, ML-based methods require training datasets. Thus, the training dataset acts as another variable. Moreover, as indicated in the data collection chapter, we consider sentiment estimation in two polarity sets, i.e. Positive and Negative (PN), and Positive, Neutral and Negative (PNN). Then the sentiment polarity is also a variable for this experiment.

In summary, the experiment involves control variables including: sentiment polarity, selected methods, the training data and the testing data.

With regard to the candidate methods, we have already selected two lexicon-based methods, and five ML-based methods, as shown in Table 3.2. In terms of data, we have common and event-based datasets for both training and testing. The details of these datasets have already been given in Table 3.4.

3.6.2 Scenario design

To explore the estimation performance under different variable values, we design a set of scenarios that combine values of those variables, shown in Table 3.5 for PN polarities and Table 3.6 for PNN polarities.

Each table consists of three sections (see Scenario column), investigating lexicon-based methods, ML-based methods trained using common-based data, and ML-based methods trained with event-based data, respectively. The Result column lists estimation results of each scenario which will be discussed in the next chapter.

3.6.3 Experimental setting

As indicated in the Research chapter, We select 1000 samples from the testing dataset for each scenario to perform 100-round testing. For ML-based methods, the training sample size is 6000 in each round, as indicated in data collection chapter.

Table 3.5: Scenario setting of sentiment estimation in polarities of Positive and Negative

Scenario	Name	Variables											Result	
		Methods							Training data		Testing data		Error	
		Lexicon-based		ML-based					Common-based	Event-based	Common-based	Event-based	Mean	Standard
		SentiStrength	SentiWordnet	NB	SGDC	LSVC	NuSVC	SVC	CA(Pos, Neg)	EA(Pos, Neg)	CA(Pos, Neg)	EA(Pos, Neg)		Deviation
Lexicon. methods	L-SS-PN-C	x							-	-	x		0.331	0.011
	L-SW-PN-C		x						-	-	x		0.407	0.013
	L-SS-PN-E	x							-	-		x	0.322	0.011
	L-SW-PN-E		x						-	-		x	0.405	0.011
ML. methods trained with common. dataset	ML-NB-PN-C-C			x					x		x		0.285	0.017
	ML-SGDC-PN-C-C				x				x		x		0.274	0.018
	ML-LSVC-PN-C-C					x			x		x		0.272	0.016
	ML-NuSVC-PN-C-C						x		x		x		0.459	0.069
	ML-SVC-PN-C-C							x	x		x		0.472	0.065
	ML-NB-PN-C-E			x					x			x	0.362	0.098
	ML-SGDC-PN-C-E				x				x			x	0.250	0.085
	ML-LSVC-PN-C-E					x			x			x	0.230	0.079
	ML-NuSVC-PN-C-E						x		x			x	0.531	0.253
	ML-SVC-PN-C-E							x	x			x	0.563	0.465
ML. methods trained with event. dataset	ML-NB-PN-E-C			x						x	x		0.491	0.013
	ML-SGDC-PN-E-C				x					x	x		0.490	0.013
	ML-LSVC-PN-E-C					x				x	x		0.500	0.014
	ML-NuSVC-PN-E-C						x			x	x		0.453	0.014
	ML-SVC-PN-E-C							x		x	x		0.505	0.015
	ML-NB-PN-E-E			x						x		x	0.184	0.011
	ML-SGDC-PN-E-E				x					x		x	0.184	0.008
	ML-LSVC-PN-E-E					x				x		x	0.177	0.010
	ML-NuSVC-PN-E-E						x			x		x	0.357	0.150
	ML-SVC-PN-E-E							x		x		x	0.174	0.011

Note: The column Result denotes experiment results. The gray cells denote smallest estimation errors in the same section. The column Name denotes a combination of candidate methods, training dataset and testing dataset, e.g.:

- L-SS-PN-C: lexicon-based method SentiStrength, in sentiment polarities of Positive and Negative (PN), tested with common-based dataset.
- ML-NB-PN-C-E: ML-based method Naive Bayes (NB), in sentiment polarities of Positive and Negative (PN), trained with common-based dataset, tested with event-based dataset.

Table 3.6: Scenario setting of sentiment estimation in polarities of Positive, Neutral and Negative

Scenario	Name	Variables											Result	
		Methods							Training data		Testing data		Error	
		Lexicon-based		ML-based					Common-based	Event-based	Common-based	Event-based	Mean	Standard
		SentiStrength	SentiWordnet	NB	SGDC	LSVC	NuSVC	SVC	CB(Pos, Neu, Neg)	EB(Pos, Neu, Neg)	CB(Pos, Neu, Neg)	EB(Pos, Neu, Neg)		Deviation
Lexicon. methods	L-SS-PNN-C	x							-	-	x		0.451	0.012
	L-SW-PNN-C		x						-	-	x		0.550	0.014
	L-SS-PNN-E	x							-	-		x	0.345	0.009
	L-SW-PNN-E		x						-	-		x	0.466	0.015
ML. methods trained with common. dataset	ML-NB-PNN-C-C			x					x		x		0.423	0.022
	ML-SGDC-PNN-C-C				x				x		x		0.414	0.016
	ML-LSVC-PNN-C-C					x			x		x		0.412	0.018
	ML-NuSVC-PNN-C-C						x		x		x		0.676	0.054
	ML-SVC-PNN-C-C							x	x		x		0.549	0.015
	ML-NB-PNN-C-E			x					x			x	0.561	0.029
	ML-SGDC-PNN-C-E				x				x			x	0.515	0.030
	ML-LSVC-PNN-C-E					x			x			x	0.526	0.026
	ML-NuSVC-PNN-C-E						x		x			x	0.364	0.026
	ML-SVC-PNN-C-E							x	x			x	0.643	0.003
ML. methods trained with event. dataset	ML-NB-PNN-E-C			x						x	x		0.678	0.012
	ML-SGDC-PNN-E-C				x					x	x		0.667	0.015
	ML-LSVC-PNN-E-C					x				x	x		0.671	0.015
	ML-NuSVC-PNN-E-C						x			x	x		0.710	0.022
	ML-SVC-PNN-E-C							x		x	x		0.708	0.015
	ML-NB-PNN-E-E			x						x		x	0.315	0.011
	ML-SGDC-PNN-E-E				x					x		x	0.309	0.010
	ML-LSVC-PNN-E-E					x				x		x	0.305	0.010
	ML-NuSVC-PNN-E-E						x			x		x	0.598	0.053
	ML-SVC-PNN-E-E							x		x		x	0.373	0.009

3.7 Sentiment Analysis: Findings of the Experiment

In this section, we show the findings and analysis of the results. It is presented and compared with and without sentiment polarity of Neutral, respectively. With in each, we start with lexicon-based methods, then ML-based methods. Finally, we compare the performance of all methods.

Table 3.5 lists sentiment estimation results with sentiment polarity of Positive and Negative (PN). With regard to lexicon-based methods, SentiStrength reaches a similar estimation error tested with both common-based and event-based data (i.e., mean error 0.331 and 0.322) which are better than SentiWordnet (i.e., mean error 0.407 and 0.405). Unexpectedly, ML-based methods, when trained with common-based dataset, and tested with the event-based dataset reach a lower minimal estimation error (i.e., mean error 0.230) than tested with the common-based dataset (i.e., mean error 0.272). The best ML-based method appears to be LinearSVC. When ML-based methods are trained with the event-based dataset, performance tests with the event-based dataset also reach a lower minimal estimation error (i.e., LinearSVC, mean error 0.177) than tested with common-based dataset (i.e. NuSVC, mean error 0.453).

Likewise, Table 3.6 shows results for the sentiment polarity Positive, Neutral and Negative (PNN). According to the results, the lexicon-based method SentiStrength reaches (again) lower estimation errors compared with SentiWordnet. In particular, it performs better with event-based testing data (i.e., mean error 0.345) than with the common testing dataset (i.e., mean error 0.451). ML-based methods shows similar patterns with the Neutral polarity as without the Neutral polarity. Specifically, when trained with common-based dataset, tests with the event-based dataset (i.e., NuSVC, mean error 0.364) perform better than tested with common-based dataset (i.e., LinearSVC, mean error 0.412). Also, when trained with an event-based dataset, this patter also holds, namely, tested with event-based data (i.e. LinearSVC, mean error 0.305) results better than tested with common-based data (i.e. SGDC, mean error 0.667.). LinearSVC reaches the lowest estimation error when trained with event and test with event dataset (i.e. mean error 0.305).

When comparing all methods, ML-based methods reach lower minimal estimation errors (tested with event-based dataset) than lexicon-based methods. In lexicon-based methods, Sentistrength reaches a lower estimation performance than SentiWordnet. For all ML-based methods, linear methods reach more consistent results. LinearSVC reaches the lowest estimation error in most scenarios, except when trained with event-based data and tested with common data, as well as trained with common data and tested with event-based data including Neutral polarity.

When comparing sentiment estimation with and without Neutral sentiment polarity, we find that all methods reach lower estimation errors without Neutral polarity than with Neutral polarity.

3.8 Discussion

We discussed the result of the sentiment analysis experiment in terms of different sentiment polarities and different training and testing datasets.

With regard to different sentiment polarities, all methods show lower sentiment estimation errors when estimating the sentiment of crowds with Positive, Negative, rather than using three polarities, i.e. Positive, Negative and Neutral. In city events, posts sent by crowds may contain more and stronger sentiments towards Positive and Negative polarities, and there may be less neutral ones than in ordinary context, which is in line with the distribution of sentiment in event-based dataset constructed in data collection section. Therefore, estimating sentiment with neutral polarity from those posts is more difficult, and consequently the estimation errors increase.

Following a similar reasoning, the lowest estimation error is reached by ML-based methods trained with event-based data and tested with event-based data with Positive and Negative sentiment polarities (i.e. LinearSVC, mean error 0.177), followed by the same method trained with common-based data and tested with event-based data with same polarities (i.e. LinearSVC, mean error 0.230). These observations may indicate that similarities between training and testing dataset in terms of content and context information may considerably affect the estimation performance. For instance, when training and testing data are both from events, even from different events, the texture characteristics, such as words, phrases, hashtags, emojis, punctuation marks, may be similar thus shows lower estimation error than training from common and tested with event-based dataset, which are less similar.

With regard to the training dataset, with sentiment polarity of Positive and Negative, the estimation error appears to be significantly distinct when ML-based methods trained with common and event-based data. This may also be explained by the similarity between the training and the testing dataset. The sentiment of posts in common-based dataset are more equally distributed, while the event-based dataset contains posts with more positive or negative sentiment. Thus, the ML-based methods trained with common-based models are less biased in sentiment estimation than trained with event-based dataset.

Lexicon-based methods tested on both common and event-based data show a similar error, which is worse than for ML-based methods: lexicon-based methods take none or limited context information (e.g. weighted lexicon-based methods) into consideration, so the estimation error is increased. This is in line with findings in Ravi & Ravi (2015) where they reviewed 161 sentiment analysis works and claim that ML-based methods result in better accuracy than lexicon-based methods because semantic orientation provides better generality. For instance, a post as ”We are having beer on the boat! #Kingsday” is identified as a neutral post by lexicon-based methods as it is interpreted as describing a fact, but it is identified as a positive post by ML-based methods in the context of the Kingsday boat parade. This may also indicate the reason why the estimation error for lexicon-based methods tested on common and event-based data is similar; the context differences between common and event scenario does not affect their decision.

3.9 Conclusion

Nowadays, city events become more and more popular. The sentiment of the crowd is one of the factors affecting crowd behaviour (Martin, 2006) and can be used to estimate and predict (negative) behaviours of the crowd, such as risky behaviours for crowd management during city events. Conventional solutions to derive such information depend on manual observations, which are expensive, prone to observation biases, and not suitable for global observations.

In this paper, we investigated the effectiveness of methods to estimate sentiment of crowds using social media text in context of city events. We created an event-based sentiment dataset consisting of social media posts from various events and major activities. Each post is annotated with sentiment polarity of Positive, Neutral and Negative using crowd sourcing. This dataset has been used for the training and testing of several methods. The main objective of our research was to investigate the performance of the candidate methods using different datasets.

We found that all candidate methods show lower estimation error with sentiment polarity of Positive and Negative, without Neutral. ML-based methods show better performance than lexicon-based methods in most situations. Specifically, the ML-based LinearSVC method reaches the minimal estimation error when trained and tested with event-based data. The findings indicate that, to predict sentiments in a crowd using social media, we recommend to use ML-based method Linear Support Vector Classifier (LinearSVC), trained with event-based data. It reaches mean estimation error at 0.177 approximately.

The results may be influenced by the construction bias of the event-based dataset, which are introduced by the various characteristics of selected events, and the unbalanced number of positive, neutral and negative social media posts. Likewise, the bias in the common-based dataset used in this research may affect the result. Moreover, in the construction of the event-based dataset, we used the Google Translation API to translate posts in other languages into English for crowd-sourcing annotation. The accuracy of translation may introduce errors in the sentiment estimation, as posts do not follow the common spell regulations.

In future work, we plan to explore more methods deriving sentiment from social media, e.g., hybrid methods that integrate the lexicon-based and ML-based methods. Also, we intend to enlarge the event-based dataset by adding more diverse events and activities, and examine the sentiment estimation performance of candidate methods across different events, or the same events in different versions. Last but not least, we will investigate the relationship between sentiment estimation performance with different sentiment schemes, e.g. a detailed sentiment scheme.

Chapter 4

Counting People in the Crowd Using Social Media Images for Crowd Management in City Events

Apart from event characteristics and sentiments, attendees' density is essential information for crowd management. To estimate the density of attendees, we first estimate the number of people (crowd size) using social media images in the city events. This answers the third research question, i.e. **RQ3. To what extent are social media images able to count people in city events?**

To this end, we first construct a social media dataset, which is used to subsequently compare the effectiveness of face recognition, object recognition, and cascaded methods for crowd size estimation. Finally, we investigate the impact of image characteristics on the performance of selected methods.

Our results show that object recognition based methods reach the highest accuracy in estimating the crowd size using social media images in city events. We also conclude that face recognition and object recognition methods are more suitable to estimate the crowd size for social media images which are taken in parallel view (i.e. the camera and the people being photographed are at a similar height), with full face selfies and in which all the people in the background have the same distance to the camera. However, so-called Cascaded methods, which count people using non-handcrafted features (learned features) applied with learning algorithms or statistical analyses, are more suitable for images taken from top view with gatherings distributed in gradient. The created social media dataset is essential for selecting image characteristics and evaluating the accuracy of people counting methods in an urban event context. It is also valuable for estimating the density of people in city events using social media.

This chapter is currently under review for journal publication.

4.1 Introduction

City events, such as sports matches, thematic carnivals and national annual festivals, are carried out in urban areas, and may attract a large number of people during a short time period. The scale and intensity of these events demand systematic approaches supporting stakeholders (e.g., event organisers, public and safety authorities) to manage the crowd. Stakeholders aim to reduce risks of incidents caused by internal and external threats, and maintain an acceptable level-of-service (Fruin, 1971; Marana et al., 1998) of the event area. The level-of-service describes the ranges of average area occupancy for a single pedestrian (Polus et al., 1983). A higher level-of-service of the event area indicates lower density of people in that area, which is safer than the lower level-of-service that refers to high density of people. Using such information together with a set of other qualitative and quantitative interpretation of the crowd, such as sentiment (Gong et al., 2019) and composition (Gong et al., 2018a), stakeholders apply predefined measures to manage the crowd. The level-of-service information can be inferred from density of people in that area, which can be further calculated using the number of people in the area and the area of the event place. Taking the national holiday, i.e. the King's Day, in the Netherlands as an example. To estimate the level-of-service in a popular attraction such as the Dam Square in the Amsterdam during the King's Day for crowd management, we can calculate the density of crowd in that area. According to Duives et al. (2015), the density of pedestrians in an area during a certain period can be defined as the number of attendees per unit area. Once the level-of-service of an area is estimated, crowd managers may apply certain measures to avoid incidents such as overcrowding in that area. Therefore, the number of people in the crowd is a valuable input for estimating the level-of-service in event area, and further for crowd management.

The crowd size can be estimated by stewards, or using crowd observation and monitoring algorithms based on data from surveys (Fang et al., 2008), cameras (Davies et al., 1995), counting systems (Daamen et al., 2016), mobile phones (Yuan, 2014; Earl et al., 2004) and public transportation systems (Luo et al., 2018; Wang et al., 2018). However, conventional methods have various disadvantages. Crowd sizes estimated by stewards and surveys contain human errors. Collecting data from cameras or counting systems can be expensive, in particular for large city events when many cameras are needed. Similar to public transportation data, sensors cannot be employed globally and may involve privacy issues. In the meantime, with the advance of technology, social media is widely used by people in city events. People on social media share their expressions by sending text and images together with timestamps and locations. Despite the drawbacks of social media data such as sparsity and dependence on individuals, images on social media may be a promising data source to estimate the size of a crowd. Figure 4.1 illustrates the relationship between the crowd size in reality and the crowd size estimates from social media images sent by the crowd during city events.

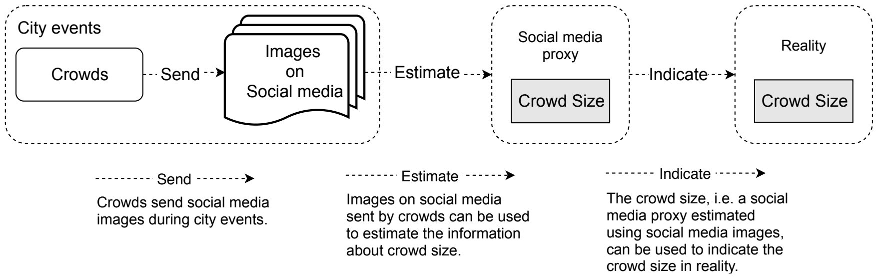


Figure 4.1: Illustration of relationship between crowd size in reality and the crowd size estimated using social media images sent by crowds in city events.

Counting people in an image is extensively studied (Chen et al., 2013; Idrees et al., 2013; Lempitsky & Zisserman, 2010; Zhang et al., 2015, 2016b). Sindagi & Patel (2018) reviewed a number of methods in their survey on estimating the number of people in the crowd from a single image. They classify these methods into two types, i.e. traditional approaches and convolutional neural network (CNN) based approaches. Saleh et al. (2015) also reviewed a set of crowd counting approaches and categorise them into direct approaches and indirect approaches. The direct approach (i.e. object based target detection) count people by identifying individual segments in the crowd and then accumulate them as the result. While the indirect approaches (e.g. pixel-based, texture-based, and corner points based analysis) count the crowd with machine learning algorithms or statistical analyses, which are considered to be more robust compared to direct methods. With respect to the context, crowd size estimation in the context of city events for crowd management differs from other contexts (e.g., shopping mall crowd monitoring (Idrees et al., 2013; Chen et al., 2013), violent behavior detection (Marsden et al., 2017)) in a set of characteristics, such as the specific topic of the event, its location, popularity, and time of occurrence. Consequently, crowd counting methods suitable for other contexts may not be fit for the context of city events. Compared to other data sources, social media images may contain more selfies and group pictures. While showing the potential of images in crowd counting analysis in general, no previous work has used social media images with diverse characteristics for crowd size estimation in the context of city events. Moreover, a dataset for which the crowd size and image characteristics are annotated for a specific context is essential in crowd counting analysis. It can be used to evaluate crowd size estimation methods and to investigate the impact of image characteristics on such results. It can also be used for machine learning methods to train their models for crowd size estimation. Recent works proposed datasets for crowd counting analysis (Chan et al., 2008; Chen et al., 2012; Idrees et al., 2013; Zhang et al., 2016a, 2015, 2016b). These datasets are diverse with respect to dense level and scene variation across images. However, none of the proposed datasets are constructed using social media images in city events, where crowd size and image characteristics have been annotated. We have identified a lack of in-depth understanding of which methods are most effective in this context, and whether their performance will be affected by the diverse image characteristics.

These research gaps lead to two research questions:

- RQ-a. Which methods are suitable to estimate crowd size using social media images in the context of city events?

- RQ-b. What is the effect of image characteristics on the accuracy of the crowd size estimation methods described in RQ-a?

To answer these research questions, we first set the scope of the crowd size to be measured in this research. Then, we selected a set of methods from diverse categories introduced by Saleh et al. (2015) to estimate the crowd size from images. In order to test the effectiveness of each method, we constructed a dataset with the number of people in the crowd annotated as well as diverse image characteristics. Next, we applied each selected method on the constructed dataset to estimate the crowd size of each image. We then analyzed the impact of image characteristics on the performance of each method. Finally, we selected the most promising method and identified under which image characteristics it is most effective.

This chapter is organised as follows. In the next section we set the scope of the crowd size to be estimated, followed by an introduction of the research methodology to examine the estimation accuracy of different methods. Then, the image characteristics from social media data are described, as well as the potential crowd counting methods. The next section introduces the social media dataset, followed by the experimental design to examine the effectiveness of the selected methods for the constructed dataset. The resulting experiment findings and corresponding analysis are then shown, followed by discussion. The paper ends with conclusions.

4.2 Definition of Crowd size

In this section, we define the crowd size that will be estimated, consisting of crowd size levels and specific numbers of people for less populated environments.

4.2.1 Crowd size levels

As indicated in the previous section, the crowd size information is essential for estimating the level-of-service in an area (Marana et al., 1998) for crowd management. The crowd size during city events is diverse, e.g. it can be small in the beginning, and become large during the peak of event activities. When the number of people is large in an image, it is difficult to get the ground truth data as manual counting becomes error-prone. Thus, in this research, we estimate the crowd size in different levels (categories) where each level corresponds a range of the number of people in an image.

Jiang et al. (2014) categorize crowd size into 5 levels, i.e. 0-10, 10-30, 30-60, 60-100, $>$100 persons. However, the number of people captured in the camera monitoring of a fixed area is far less diverse than the number of people in social media images, e.g. social media selfies may only contain a few people, while some panorama pictures may contain a large number of people in a square or on the street in city events. Therefore, we adjusted the categories of crowd size used by Jiang et al. (2014) into a larger scope.

We performed a prior investigation on the pilot social media dataset, introduced later. We found that around 30% of the images do not contain people, and around 50% of the images contain less than 20 people, most of which are selfies and group pictures. Therefore, we set the first crowd size level to contain no people in images, denoted as 0, and the second level containing a number of people between 1 and 20, denoted as 1. Furthermore, we define 3 levels with a number of people above 20. Though the scoping of levels 2, 3, and 4 is different from the categories used in Jiang et al. (2014), this difference of scoping has less impact on this research as the proportion of images containing more than 20 people is less than 20% in social media images.

We summarise the scoping of crowd size level as follows:

- Level 0 denotes the number of people is 0, i.e., no people.
- Level 1 denotes the number of people between 1 and 20.
- Level 2 denotes the number of people between 20 to 100.
- Level 3 denotes the number of people between 100 to 250.
- Level 4 denotes more than 250 persons.

4.2.2 Number of people in the crowd in less populated images

In a less populated environment, where the number of people is less than 20 (level 1), it is possible to count the specific number of people in an image and derive the ground truth data. Thus, exploring the accuracy of the methods for level 1 can be more precise. Therefore, in this research, we also estimate the specific number of people in less populated images, i.e. images with crowd size level 1.

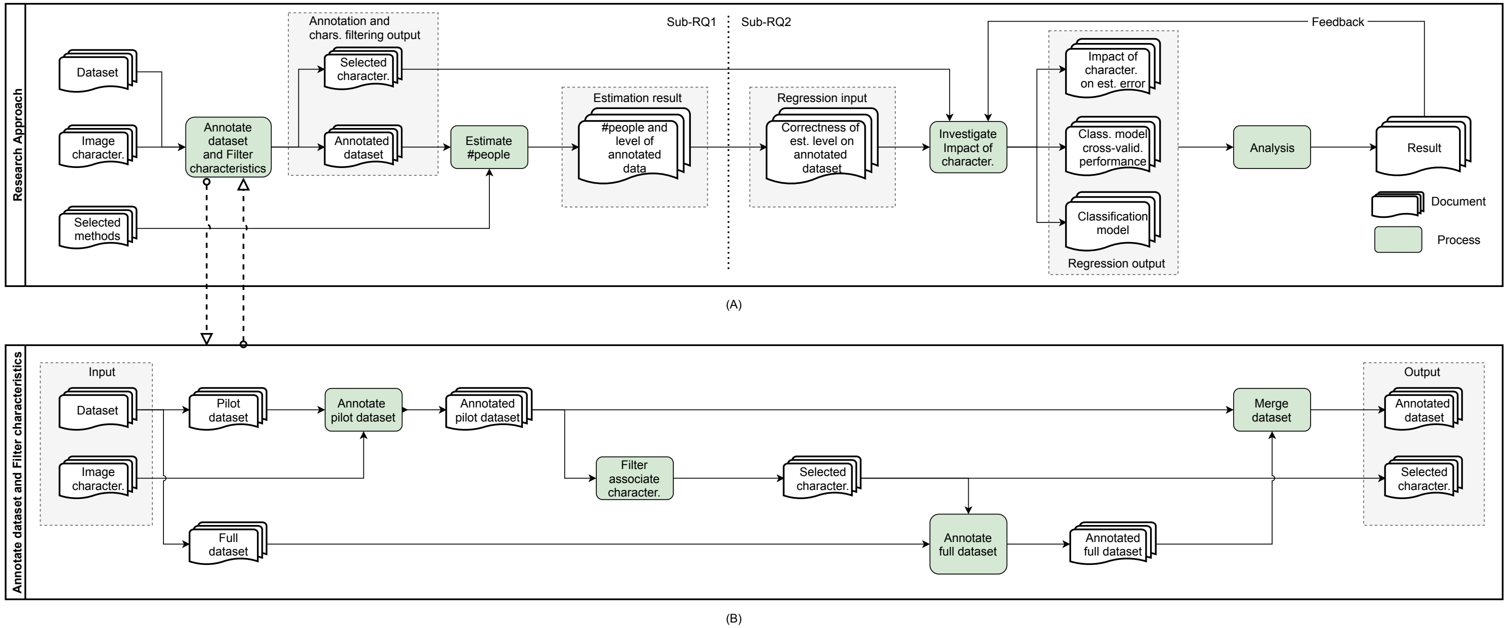


Figure 4.2: The process of exploring the effectiveness of the selected methods on crowd size estimation using social media images in city events, and investigating the impact of image characteristics on the estimation results. The symbols in green denote process steps. The symbols in grey denote input and output.

4.3 Research approach

In this chapter, we introduce the research approach and assessment metrics for answering the research questions. We illustrate each step of the research approach shown in Figure 4.2.

To investigate the first research question, i.e. the effectiveness of methods in crowd size estimation using social media images in city events, we select a set of crowd counting methods with diverse techniques reviewed in surveys (Sindagi & Patel, 2018; Saleh et al., 2015). To estimate the crowd size using selected methods, we construct a social media dataset of images that are collected during city events considering various aspects, such as event topics, editions, length of duration, cities and area in the city. The selected methods are applied on the annotated dataset, yielding the estimated crowd size for each image. The accuracy of crowd size estimation is calculated by comparing the estimation result to the ground truth, which indicates whether the crowd size level of an image is the same as the ground truth.

To answer the second research question, i.e. investigate the impact of image characteristics on the crowd size estimation of different methods, we generated a set of image characteristics from both the crowd management perspective and the social media image perspective. Image characteristics generated from crowd management include conditions such as indoor or outdoor, and the urban environment where pictures are captured such as square, street, canal, and park. Image characteristics generated from social media image perspective consist of characteristics may affect crowd counting effectiveness, such as people_present, view, and selfie_face.

We zoom in to the dataset annotation and image characteristics filtering in Figure 4.2 (B). As the image characteristics generated in the previous step may contain highly correlated characteristics, we perform a characteristic selection procedure to filter out high correlated image characteristics. To do so, we randomly selected a set of images from the total dataset as a pilot dataset and annotated these with values of images characteristics and the crowd size as ground truth. After checking the correlation on all characteristics annotated in the pilot dataset, the least correlated image characteristics are screened out. The full dataset is then annotated with the selected image characteristics and the crowd size. The output of this sub-process is the annotated dataset merging the Pilot dataset and the Full dataset, named Total dataset.

To investigate the impact of image characteristics on the accuracy of the selected methods, we train a classifier using a logistic regression algorithm for machine learning (Dreiseitl & Ohno-Machado, 2002) for each method on the value of image characteristics with the correctness of the crowd size level estimation. This step outputs a set of impacts (coefficient) for each image characteristic. It also produces a classification model as a by-product with the average performance of the model calculated from the cross-validation. We analyse these outputs and provide it as a feedback to improve the investigation of the impact of image characteristics on effectiveness of methods in crowd size estimation.

4.3.1 Comparison metrics

We use a set of comparison metrics to analyse the accuracy of methods in crowd size estimation and the impact of images characteristics on the estimation result.

The estimation accuracy of methods

The estimation performance is assessed using the estimation accuracy, which is calculated for each method i. The estimation error A_i is calculated by the amount of correctly identified images M_i^{true} divided by the total sample size M_i, see Eq. 4.1. We are aware of the drawbacks of this measure with respect to the (hard) boundaries of the crowd size levels, e.g. assuming an image contains 99 people, i.e. ground truth crowd size level 2, while the estimated number of people is 101, i.e. crowd size level 3. Though the difference between ground truth value and the estimated number of people in the image is small, the estimated crowd size level is incorrect, which seems to be an overreaction. To compensate this, we also check the ground truth and estimation close distance in adjacent levels.

To explore the insights of the estimation performance of different methods, we further show the distribution of estimated crowd size levels compared with the ground truth in Table 4.7, and the distribution of the estimated number of people in crowd size level 1 compared with the ground truth in Figure 4.6.

$$A_i = \frac{M_i^{true}}{M_i} \tag{4.1}$$

Classification performance

The investigation of the impact of image characteristics on the accuracy of the methods in crowd size estimation outputs the impact of each image characteristic, a classification model, and the cross-validation performance of this classification model. The performance of the classification model is indicated by the cross-validation performance. This cross-validation performance of the classification model is measured with metrics for binary classification, i.e. *Precision*, *Recall* and *F1_Score* (Powers, 2011). *Precision* refers to the percentage of classified results are correct among all classified results, while the *Recall* refers to percentage of correct items have been classified among all correct items. The *F1_Score* is simply the harmonic mean of the precision and the recall.

4.4 Social media image characteristics

In this section, we identify a set of scene characteristics of the images (in the following referred to as Image Characteristics) to investigate their impact on the accuracy of crowd counting methods. The image characteristics consist of requirements from crowd management such as indoor/outdoor and urban environment shown in each image (as the purpose of this research is to provide information about crowd size for crowd management), and characteristics of images posted from social media in terms of, e.g. image type (selfie or group picture) and distribution of crowds, that may affect the performance of crowd counting methods. The image characteristics are further categorised into three types, i.e. global, frontend and backend image characteristics, which will be introduced in the following sections. The detailed definitions of image characteristics are given in Table 4.1, and corresponding examples are shown in Figure 4.3.

Table 4.1: Identified image characteristics

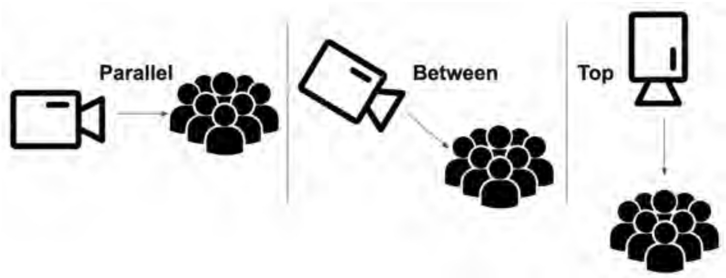
Perspective	Category of image characteristic	Image characteristic	Definition	Values and example image(s)
Crowd management	Global	Condition	Whether this photo is taken indoor or outdoor?	- Indoor: the picture is captured indoor, e.g., Figure 3a. - Outdoor: the picture is captured outdoor, e.g., Figure 3b.
	Global	Urban environment	The place where the picture is captured, such as square, street, canal, park and others.	- Square: e.g., Figure 3d, 3e. - Street: e.g., Figure 3c. - Canal: e.g., Figure 3h. - Park: e.g., Figure 3b. - Others: e.g., Figure 3a, 3f, 3g, 3i.
Social media	Global	People present	Whether this image contain people.	- Yes: the picture contains people, e.g., Figure 3a, 3b, 3c. - No: the picture contain no people, e.g., Figure 3c.
	Global	View	The view of the camera to people, i.e. top, parallel, or between top and parallel.	- Top: the people in the picture is captured from top, e.g., Figure 3a, 3d. - Parallel: the people in the picture is captured in the same level with the camera, e.g., Figure 3b, 3f. - Between: the people in the picture is captured between top and parallel, e.g., Figure 3e.
	Front.	Has selfie	Whether this image contain selfie.	- Yes: there are selfie people in the frontend of the picture, e.g., Figure 3b, 3f. - No: there is no selfie people in the frontend of the picture, e.g., Figure 3a, 3e.
	Front.	Selfie face	The different types of faces captured in this image, such as full, partial, back, mixed, or none face.	- Full: full face is shown in the picture, e.g., Figure 3f. - Partial: only part of the face is shown, e.g., Figure 3g. - Back: the back face or hack head is shown, e.g., Figure 3h. - Mixed: mixed faces, e.g., Figure 3b. - None: such as only part of the body is shown, e.g., Figure 3i.
	Back.	Has gatherings	Whether there are people in the backend of the picture.	- Yes: there are people in the backend of the picture, e.g., Figure 3e, 3j. - No: there is no people in the backend of the picture, e.g., Figure 3b, 3f, 3h, 3i.
	Back.	Gatherings distribution	The distribution of the gatherings in the backend of the picture, i.e. fixed, gradient.	- Fixed: gatherings in the backend of the picture are in the same distance from the camera, e.g., Figure 3j, 3m. - Gradient: gatherings in the backend of the picture are gradually far away from the camera, e.g., Figure 3d, 3e, 3l.
	Back.	Gatherings clarity level	The clarity level of people who is the most clearest one in the gatherings in respect to be identified.	- A: Very clear. The face of each person is shown clearly. The detail of features on the face can be identified, e.g., Figure 3f. - B: Clear, the face of each person is clear. The features on the face can be identified, but the detail of features can not be identified, e.g., Figure 3a, 3b, 3g, 3h. - C: Clear. The face of each person is still observable but the features on the face is not distinct. The shape of each person is clear, e.g., Figure 3e. - D: Less clear. Each person is only shown as a shape. The face is not observable, e.g., Figure 3j. - E: Very unclear. Each person is shown as a dot. Their face and shape is totally unobservable, e.g., Figure 3d.

***Frontend and backend** there are two layers in the picture where people are present, that is, in the Frontend layer and in the Backend layer, illustrated in Figure 4.4(b).

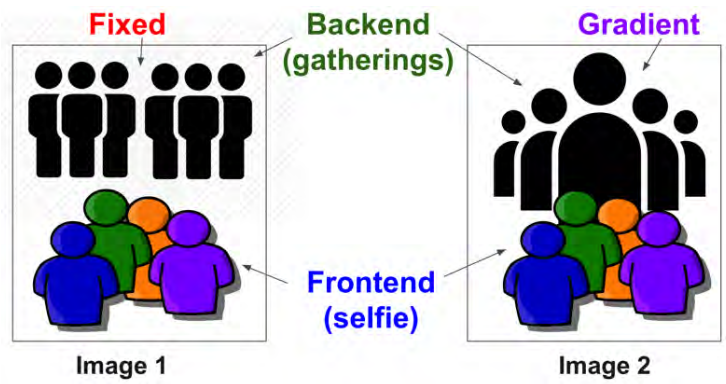
Figure 4.3: Sample of images with diverse characteristics, as listed in Table 1. These images are collected from social media in city events of King's Day 2016-2018, and Europride 2017, both in Amsterdam, the Netherlands.

4.4.1 Crowd management perspective characteristics

As in this research, we investigate the effectiveness of methods counting people for crowd management in city events carried out in an open space (denoted as outdoor), such as in the street, sports stadium, and conference centre, rather than in a closed space (denoted as indoor) such as in a room or shop. Thus we identify images if they are taken indoor or outdoor. In the meantime, crowd managers typically apply predefined measures to manage the crowd. The confinedness of an area where the crowd is located is relevant to select measures; do people have an area to go? Is there water or



(a) View



(b) Frontend, backend layers of an image, and Gatherings_distribution

(c) Gatherings_clarity_level

Figure 4.4: The schematic interpretation of image characteristics in terms of views, has_gatherings, gatherings_distribution, and gatherings_clarity_level, listed in Table 4.1. Social media images are collected from social media in city events of King's Day 2016-2018, and Europride 2017, in Amsterdam, the Netherlands.

a railway track or another inaccessible area around? Therefore, we need to identify the urban environment shown in the images, such as square, street, canal, park, which may affect the effectiveness of methods counting people in the crowd. We categorise these two image characteristics as global characteristics because they exist in all images.

4.4.2 Social media perspective characteristics

To identify characteristics on social media images that may affect the effectiveness of counting people methods, we reviewed social media images from the pilot dataset (which will be introduced in the section Data Collection). As all selected methods for counting people (which will be introduced in the section Selection of Crowd Size Estimation Methods) are based on the content of images, we also focus on characteristics of images content rather than other information of images, such as meta-data. Firstly, we found that in only few images no people are present, which makes the social media images indeed a good resource to count people. Among all images with people present, people are captured by the camera from different viewpoints, such as Top, Parallel, and between Top and Parallel, illustrated in Figure 4.4(a), which may affect the effectiveness of methods for counting people through identifying shapes and faces. These two image characteristics (whether people present and camera viewpoint) are also categorised as global image characteristics as they exist in all images. In addition, we found that people are present in the two layers in the picture, i.e. the frontend layer which close to the camera lens, and the backend layer far away from the camera lens, as illustrated in Figure 4.4(b). Images with people in the frontend are normally selfies. People in the backend of an image are so-called gatherings (denoted as gatherings). The different size and shapes of people in the frontend and backend may affect effectiveness of people counting methods. Based on the viewpoints and layers where the people are captured, images containing people can be further categorized into three types, i.e. only selfie, selfie with gatherings, and only gatherings. For selfies containing people in the frontend, the face of people may be diverse, e.g. Full face, Partial face, Blocked face, Back face (or Hack head), or No face (i.e. only showing a body rather than a face). This may affect faces-based methods in counting people. For images containing gatherings, the distribution of gatherings can be divided into two types, i.e. Fixed and Gradient, illustrated in Figure 4.4(b). Fixed gatherings indicate that the people have a similar distance to the camera. Gradient gatherings, on the contrary, have different distances to the camera, with people shown in a smaller size having a longer distance. The different distributions of gatherings lead to different size and clarity of faces and shapes, which may affect the effectiveness of people counting methods. Among all images with gatherings in the backend, the clarity of people is different: some images are quite blurry, while others are very clear. The different clarity of gatherings in the backend may affect the effectiveness of methods which count people based on faces and textures (e.g. cascaded methods). We then categorize images with people in gatherings into different levels in terms of clarity, illustrated in Figure 4.4(c), ranging from A to E, where A indicates the highest clarity and E indicates the lowest clarity. In clarity level A, the faces and detail feature on faces are clear and can be identified. In level B, the face is clear. The features on the faces are observable but can not be identified. In level C, only the faces are observable while the features on the face are not distinct. The shape of people is clear. In level D, each person is only shown as a shape. In level E, each person is shown as a dot. The detail rules and examples for distinguishing the levels of clarity are listed in the row of Gathering clarity level in Table 4.1. Further, We categorise image characteristics of Has selfie and Selfie face types as frontend image characteristics as they exist in the frontend layer of images. The image characteristics of Has gatherings, Gatherings distribution, and Gatherings clarity level are categorised as backend image characteristics, because they exist in the backend layer of images.

4.5 Selection of crowd size estimation methods

In this section we select the methods to perform crowd size estimation on social media images. As indicated in the introduction, there is no existing literature comparing the performance of crowd size estimation methods using social media data in the context of city events. However, counting the number of people in an image is not a novel problem. Many works discussed this topic and proposed methods to solve this problem, see (Chen et al., 2013; Idrees et al., 2013; Lempitsky & Zisserman, 2010; Zhang et al., 2015, 2016b). More than 60 methods are reviewed in surveys about counting people from images (Sindagi & Patel, 2018; Saleh et al., 2015; Ryan et al., 2015). As introduced in Introduction Section, these methods can be categorized with respect to different approaches, i.e. direct approaches and indirect approaches. The methods using direct approaches identify persons using handcrafted features in an image and accumulate them as the number of people in an image. The handcrafted features refer to properties derived

Table 4.2: Selected methods for counting people in the crowd

Approach	Approach category	Name	Features	Performance
Face recognition	Direct methods	Faceplusplus	Face of people	99.50% accuracy in Wild (LFW) test
Object recognition		Darknet Yolo	Shape of people	Outperformed related methods
Convolutional neural	Indirect methods	Cascaded A	Learned features	Lowest MAE in random pictures about city events
network machine learning		Cascaded B	Learned features	Lowest MAE in pictures of busy streets in city events˜

*MAE: Mean absolute error.

beforehand by human experts using the information present in the image itself, such as the face, head, shoulder and legs of people (Nanni et al., 2017). The methods with indirect approaches count people using non-handcrafted features applied with learning algorithms or statistical analyses. The non-handcrafted features are also called learned features, which is learned by machine learning algorithms using data rather than handcrafted features.

In this research we select several methods to count people from images. The selection criteria are as follows: the selected methods should be 1) diverse in mechanism, 2) diverse in specific features used for identifying and counting people, and 3) should have a high performance in comparison to related methods. To meet the first criterion, we select both direct and indirect methods. For direct methods, we further consider methods with different features for identifying and counting people, such as faces and objects. We select Faceplusplus (Zhou et al., 2015) and Darknet Yolo (Redmon & Farhadi, 2017), which identify people through face recognition and object recognition, respectively, and reach high performance (Zhou et al., 2015; Redmon & Farhadi, 2017). The selected indirect methods are convolutional neural network based Cascaded methods (Sindagi & Patel, 2017) with version A and B, which reach significantly better results in comparison with related methods (Sindagi & Patel, 2017). The details of the selected methods are listed in Table 4.2 and described as follows.

Face Recognition: Faceplusplus (Face++)

Faceplusplus (Face++) is a method widely used for identifying people by their faces (Zhou et al., 2015). The face recognition model is established based on a deep convolutional neural network which is trained with 5 million labelled faces with about 20,000 individuals. It reaches 99.50% accuracy in the test of recognizing faces in the database Labelled Faces in the Wild (LFW) (Zhou et al., 2015), a database of faces designed for studying the problem of unconstrained face recognition. This method detects faces in each image and provides the amount of faces in each image, based on which we can calculate the crowd size of each image.

Object Recognition: Darknet Yolo (You Only Look Once)

Darknet Yolo (You Only Look Once) is a state-of-the-art, neural network based machine learning method for real time object detection (Redmon et al., 2016; Redmon & Farhadi, 2017). It is constructed based on the Darknet, an open source neural network framework (https://pjreddie.com/darknet/). It can recognise a large number of objects including persons and reaches a mean average precision of 78.6 on the PASCAL Visual Object Classes Challenge 2007 (PASCAL VOC 2007), which outperformed other algorithms widely used in this fields such as Fast R-CNN, SSD300 and SSD500 (Redmon & Farhadi, 2017). This method detects people and exports the number of people in each image.

CNN-based: Cascaded methods

The cascaded method is a state-of-the-art, convolutional neural network based (CNN-based) machine learning method for estimating the number of people in a high density context in an image (Sindagi & Patel, 2017). It consists of two trained models, i.e. Cascaded A and Cascaded B. Both models are trained with different parts of the Shanghai Tech dataset (Zhang et al., 2016b) which contains 1,198 annotated images with a total of 330,165 people. Cascaded A is trained with part of images in the Shanghai Tech dataset which are randomly crawled from Internet about the Shanghai Tech event and most of them have a large number of people. The Cascaded B is trained with part of images in the Shanghai Tech dataset which are taken from busy streets of metropolitan areas in Shanghai during the Shanghai Tech event (Sindagi & Patel, 2017; Zhang et al., 2016b). According to the comparison (Sindagi & Patel, 2017), the performance of both cascaded models outperformed popular methods used in this fields such as MCNN (Zhang et al., 2016b), Idrees (Idrees et al., 2013), Walach & Wolf (2016) and Zhang et al. (2015). The Cascaded methods estimate number of people in each image. Then, for each image the crowd size level corresponding to this number of people is assigned.

4.6 Data collection and annotation

In this section, we construct a dataset containing annotated social media images in city events, for deriving ground truth to investigate the effectiveness of different methods in crowd size estimation, as well as impact of image characteristics on this effectiveness. To collect these social media images, we first select a set of events and activities during these events. Then, social media images taken during these events and activities are collected from Instagram, the most popular image based social network (Yang et al., 2016; Gong et al., 2018a). After collecting the data, we use these social media images to derive the ground truth (annotated dataset). As we also want to investigate the impact of image characteristics, these images are also used for identifying image characteristics. We introduce each step in detail in the following.

4.6.1 Event selection

Social media data are collected from various city events and the activities during these events. To avoid bias in selecting city events as well as activities, We identify the requirements regarding the events and activities considering diversity in terms of cities, event characteristics, and their major activities. The selected events are listed in Table 4.3. It contains 4 different city events with different topics, i.e. King's Day is a celebration of the King's birthday in the Netherlands, Europride is a LGBT festival, Sail is a nautical event, and Feyenoord represents the Feyenoord football fan riots in 2017. We have collected data during three editions of King's Day event, in 2016, 2017 and 2018. The selected events are diverse in duration, ranging from less than 1 day to 9 days. They are also diverse in cities and areas in the city, e.g. while the Feyenoord event is in Rotterdam, all other events take place in the city of Amsterdam. Except for the Sail event and the Feyenoord event, which took place in the IJ area (bay area) and around the football stadium, respectively, the events took place in the city centre.

4.6.2 Social media data collection

Instagram, the image based social media network, is widely used by people to share pictures (Yang et al., 2016; Gong et al., 2018a). Therefore, we collect images from Instagram to construct the social media dataset. Instagram

Table 4.3: City events and their characteristics for constructing social media dataset

Name	Year	City	Area	Date	Term	Topic
King's Day	2016	Amsterdam	City center	27-04-2016	1 day	King's birthday celebration
	2017	Amsterdam	City center	27-04-2017	1 day	King's birthday celebration
	2018	Amsterdam	City center	27-04-2018	1 day	King's birthday celebration
Europride	2017	Amsterdam	City center	29-07-2017 to 06-08-2017	9 days	LGBT* festival
Sail	2015	Amsterdam	IJ (bay) area	19-08-2015 to 23-08-2015	5 days	Nautical event
Feyenoord	2017	Rotterdam	Around stadium	07-05-2017 after 14:30	Less than 1 day	Football fan riots

LGBT stands for lesbian, gay, bisexual, and transgender.

images are collected through the API of Instagram platforms using SocialGlass (http://social-glass.tudelft.nl/), an integrated system for collecting and processing social media data (Bocconi et al., 2015; Psyllidis et al., 2015a). In the end, we collected 2,028 Instagram images sent during selected events.

4.6.3 Data annotation

The total dataset is split randomly into two sub-datasets, i.e. pilot dataset and full dataset, each containing around 50% of Instagram images collected during selected events. The pilot dataset is used for identifying and selecting image characteristics, while the full dataset will be later annotated with the image characteristics selected from the pilot dataset, and further merged with the pilot dataset to derive ground truth and investigate the impact of image characteristics on crowd size estimation. Table 4.4 and 4.5 lists composition of two datasets in terms of crowd levels and images characteristics.

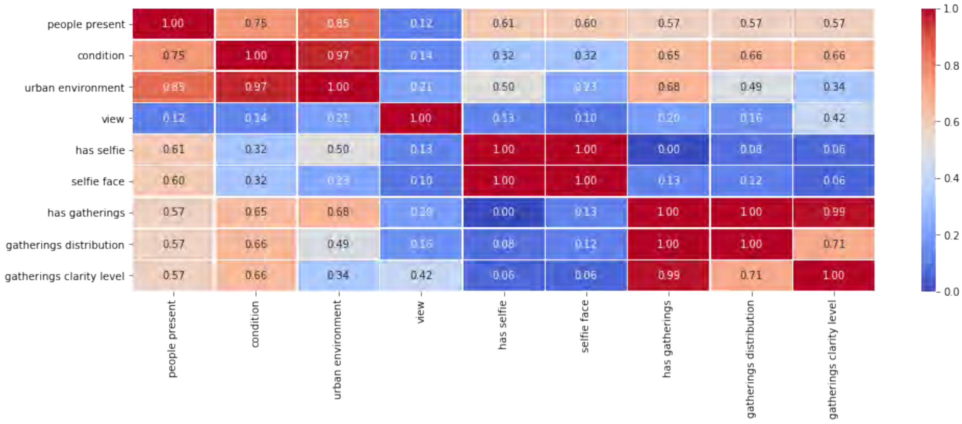
Pilot data annotation

The pilot dataset is manually annotated with regard to the crowd size as the ground truth, and values of identified image characteristics in section of Social media image characteristics.

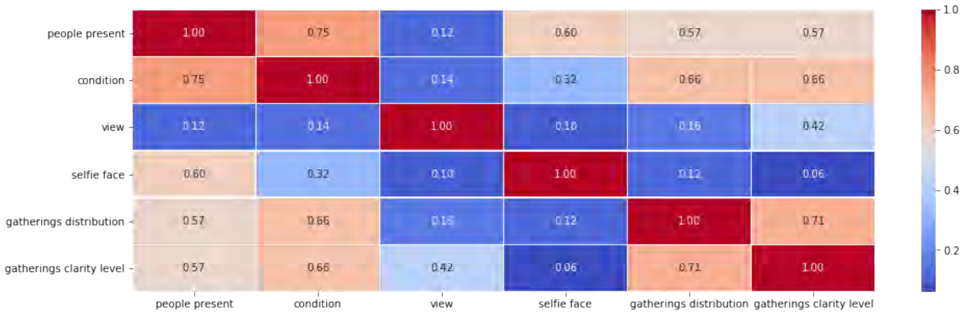
Characteristics selection

As the image characteristics are identified from two different perspectives, i.e. the crowd management and social media. There might be overlaps, i.e. strong associated characteristics, exist between these two perspectives. To analyse the impact of image characteristics on the crowd size estimation accuracy, it is necessary to identify the associated characteristics and select the representative one.

We employ Cramer's V (Cramér, 1999) to calculate the effect size of an association between each pair of characteristics. Cramer's V varies between 0 and 1. A value close to 0 shows little association between two characteristics. However, a value close to 1 indicates a strong association. Heatmaps in Figure 4.5(a) shows the strength of association between each pair of characteristics. Redder cells denote strong associations while bluer ones denote weaker associations.



(a) Identified image characteristics



(b) Selected image characteristics

Figure 4.5: The association heatmap of identified image characteristics, calculated by Cramer's V based on annotated pilot dataset.

According to Figure 4.5(a), strong associations are observed between condition and 'urban environment', 'has selfie' and 'selfie face', 'has gatherings' and 'gatherings distribution'. For each above pair, one image characteristic will be removed, if:

- The association with other characteristics is higher than another characteristic.
- It is not required by crowd management, listed in Table 4.1.

For the pair 'condition' and 'urban environment', the 'condition' characteristic indicating an indoor or outdoor environment is key information for crowd management (Martella et al., 2017). It is more important than a specific location as is indicated by the 'urban environment' characteristic. Thus, the 'condition' characteristic will be kept. For the other two pairs, i.e. 'has selfie' with 'selfie face', and 'has gatherings' with 'gatherings distribution', the latter characteristics in each pair, i.e. 'selfie face' and 'gatherings distribution', contain richer information than the former ones, and has less association in average with other image characteristics. Consequently, the latter ones are kept.

Thus, the following image characteristics will be removed: 'urban environment', 'has selfie', and 'has gatherings'. The association heatmap of the selected image characteristics is shown in Figure 4.5(b).

Full dataset annotation

After selecting the image characteristics, we annotated the selected image characteristics and the crowd size (ground truth) for the rest of the dataset (Full dataset). For this annotation we have used crowd-sourcing (Schenk et al., 2009): the selected image characteristics and crowd size of each image are determined by multiple people and the majority judgement is taken as the ground truth. We performed the crowd-sourcing operation using Figure Eight (https://www.figure-eight.com/), a popular crowd-sourcing platform where data annotation tasks are distributed to a large number of people and the majority annotation is automatically calculated.

After the annotation of the full dataset, it is merged with the annotated pilot dataset to come up with the annotated total dataset.

Dataset descriptive statistics

The descriptive statistics of the dataset in terms of selected image characteristics are listed in Table 4.4 and 4.5, with highlights for the largest proportions in the Total dataset in terms of crowd level or categories of image characteristics. The total dataset contains 2,028 Instagram images, of which the pilot dataset contained around 47.14% and rest are part of the full dataset. The crowd level distributions among all social media images show similar pattern across datasets, i.e. almost one third of them contain no people, and half of them contain number of people less than 20. While, Around 12% of them contain number of people between 20 to 100. Images containing more than 100 people are rare.

With regard to the image characteristics in both datasets, among more than two third of images in which people are present, 74% of them in average are taken in the outdoor, and around 89% of them are taken in the parallel view. With regard to the frontend image characteristics, about 63% of images which contain people are selfie photos, and 79% of such selfies captured the full face of people. With regard to the backend image characteristics, among all images containing gatherings, around 76% of them the gatherings are shown in gradient distribution, i.e. the gatherings are gradually far away from the camera. In around 36% and 27% of such cases, clarity levels are B and C, implying that we can see the face without the detailed features of the faces of those people in the gatherings, or we can only see their shapes, respectively.

Table 4.4: Descriptive statistics of the annotated social media image dataset in terms of crowd level and categories of image characteristics, part 1

Dataset		Crowd level						Global image characteristic									
Name	# images~	% of each dataset						People present, % of each dataset			Condition, % of people present			View, % of people present			
		0	1	2	3	4		Yes	No		Indoor	Outdoor		Top	Parallel	Between	
Pilot	956	30.45%	51.93%	13.44%	2.65%	1.53%	100.00%	69.55%	30.45%	100.00%	16.69%	83.31%	100.00%	1.02%	94.58%	4.39%	100.00%
Full	1,072	32.70%	48.32%	11.81%	4.72%	2.45%	100.00%	67.30%	32.70%	100.00%	21.59%	64.91%	100.00%	3.37%	84.08%	12.55%	100.00%
Total	2,028	31.64%	50.02%	12.58%	3.74%	2.02%	100.00%	68.36%	31.64%	100.00%	19.24%	73.74%	100.00%	2.25%	89.12%	8.64%	100.00%

The cell in **grey** denotes it takes the largest proportion in the Total dataset in terms of the crowd level or categories of image characteristics.
The **0, 1, 2, 3, 4** denote the crowd level.
The cell valued **73.74%** in column **Outdoor** in **"Condition, % of images people present"** denotes the 73.74% of people present images were taken outdoors.

Table 4.5: Descriptive statistics of the annotated social media image dataset in terms of crowd level and categories of image characteristics, part 2

Dataset		Frontend image characteristic							Backend image characteristics									
Name	% of total dataset	Selfie, % of people P.	Selfie face, % of selfie images						Gathering., % of people P.	Gatherings Distribution, % of images contain Gathering			Gatherings clarity level, % of images contain Gatherings					
			Full	Part	Back	Mixed	None	Fixed		Fixed	Gradient		A	B	C	D	E	
Pilot	47.14%	66.03%	75.83%	9.53%	4.66%	9.09%	0.89%	100.00%	62.37%	21.36%	78.64%	100.00%	23.24%	39.67%	26.76%	9.86%	0.47%	100.00%
Full	52.86%	60.19%	81.39%	9.64%	2.47%	2.91%	3.59%	100.00%	59.51%	25.62%	74.38%	100.00%	10.43%	32.65%	28.12%	23.58%	5.22%	100.00%
Total	100.00%	62.99%	78.60%	9.59%	3.57%	6.02%	2.23%	100.00%	60.88%	23.53%	76.47%	100.00%	16.72%	36.10%	27.45%	16.84%	2.88%	100.00%

The **grey** cell denotes it takes the largest proportion in the Total dataset in terms of categories of image characteristics.
The **A, B, C, D, E** denote the gatherings clarity level.

4.7 Experimental setup

We set up experiments to perform crowd counting analysis using social media images in city events to answer two research questions (RQ-a and RQ-b) of this study. To answer the first research question (RQ-a) we set up an experiment to study the crowd size estimation accuracy of different methods. To answer the second research question (RQ-b) we set up an experiment to investigate the impact of selected image characteristics on the correctness of crowd size estimation by each method. In the following subsections, we introduce each experiment in terms of their variables, expected results and process.

4.7.1 Experiment 1: Crowd size estimation accuracy

We set up the first experiment to assess the estimation accuracy for each selected method to estimate the crowd size. The independent variable in this experiment is the accuracy of the selected methods and the annotated social media images in the dataset. The dependent variable is the crowd size of each image estimated by the each method. In addition to the estimated crowd size, this experiment also outputs the measures A_i defined in Eq.4.1 for comparing the effectiveness of different methods. The experiment process is listed as follows:

- For each method, we perform crowd size estimation on social media images in the dataset, yielding a set of crowd size estimated for each image.
- Calculate the measures A_i defined in Eq.4.1 on the set of estimated crowd size for each method.

4.7.2 Experiment 2: Impact of image characteristics on crowd size estimation

The second experiment is set up to investigate the impact of image characteristics on the accuracy of the estimation of crowd size level estimation by the different methods. The independent variables consist of the selected image characteristics and the correctness of crowd size level estimated by each method. The dependent variable is the impact (coefficient) of each image characteristic on the crowd size level correctness for each method.

After performing the experiment process, in addition to the image characteristics impact, the experiment also outputs a classification model with cross-validation performance as a side product for each method. The classification model classify input images into bi-categories, i.e. whether the crowd size in an image can be correctly estimated, while taking into account the image characteristics. For instance, the crowd size in a selfie image containing people with full faces, captured in parallel view without mass gatherings in the backend may be correctly estimated by Faceplusplus or Darknet Yolo methods rather than Cascaded methods. The process of this experiment is listed as follows:

- Calculate whether the estimated crowd size level is the same with the ground truth for each image, and set the result as the dependent bi-categorical variable.
- Train a binary classifier using logistic regression algorithm for machine learning (Dreiseitl & Ohno-Machado, 2002) for each method with image characteristics and the estimation correctness by this method. To assess how the classifier will generalize to an independent dataset (Kohavi et al., 1995), we apply 5-folds cross-validation (Guyon, 1997) in the training process.
- Record the impact (coefficient) of each selected image characteristic from the trained model and measure the classification performance from cross-validation by *Precision*, *Recall* and *F1_Score* (Powers, 2011) introduced in the previous chapter.

4.8 Crowd counting analysis: findings of the experiment on crowd size estimation and image characteristics impacts

In this section, we analyse the accuracy of the estimated crowd size for each method, as well as the impact of image characteristics on the crowd size level estimation for each method listed in Table 4.2.

4.8.1 Crowd size estimation from social media images in city events

Table 4.6 lists the results of the crowd size estimation of different methods using social media data in city events. Here, we distinguish different levels of estimation. When the estimated level is 1 (so less than 20 persons), we also specify the estimated number of people.

Crowd size level estimation

According to Table 4.6, Faceplusplus (65.00%) and Darknet Yolo (72.01%) reach 2-3 times higher accuracy than Cascaded A (24.72%) and Cascaded B (35.43%). Faceplusplus and Darknet Yolo underestimate the crowd size in a large number of images, while Cascaded A and B predict too high values. As Faceplusplus and Darknet Yolo count people by identifying their faces or shapes, the crowd size in dense images is underestimated, as the faces and shapes might not be available in this type of images.

Table 4.6: Crowd size estimation by different methods

	Crowd size level estimation				# People in crowd size Level 1 estimation			
	# Underestimate	# Accurate est.	# Overestimate	Est. Accuracy	# Underestimate	# Accurate est.	# Overestimate	Est. Accuracy
Faceplusplus	679	1,318	31	65.00%	635	381	8	37.21%
Darknet Yolo	531	1,460	37	72.01%	515	390	119	38.09%
Cascaded A	30	501	1,496	24.72%	371	21	632	2.05%
Cascaded B	65	719	1,244	35.43%	305	50	669	4.88%

The **grey** cell denotes model with the highest estimation accuracy.

To compare the estimated levels with the ground truth, we show the distribution of estimated levels for each method in Table 4.7. The diagonal of the table shows the percentage of images that are correctly estimated by each method. According to the table, Faceplusplus and Darknet Yolo produce higher percentage of correct estimation in less dense levels 0 and 1. Instead, the Cascaded A and B produce more correct estimations for higher levels. This may also be caused by the distinct features used by different methods in detecting people. We can thus conclude that Faceplusplus and Darknet Yolo are more feasible in low dense environments, while Cascaded A and B are fit for high dense environments. While comparing the Faceplusplus and Darknet Yolo, the latter method reaches better accuracy than the former one; which may indicate that in social media images, even in low dense environment, shapes are more available or valuable than faces to be

Table 4.7: Crowd size level estimation by different methods

Method		Ground Truth					
	Levels	0	1	2	3	4	Total
Face ++	0	95.00%	30.90%	45.60%	55.10%	59.50%	54.00%
	1	5.00%	69.00%	53.60%	43.60%	40.50%	45.80%
	2	0.00%	0.10%	0.80%	1.30%	0.00%	0.20%
	3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Darknet	0	94.40%	16.40%	24.50%	19.20%	16.70%	41.50%
Yolo	1	5.60%	83.50%	74.30%	78.20%	81.00%	58.20%
	2	0.00%	0.10%	1.10%	2.60%	2.40%	0.30%
	3	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Cascaded A	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	1	31.70%	32.40%	4.20%	5.10%	0.00%	26.80%
	2	35.90%	39.80%	40.20%	7.70%	9.50%	36.80%
	3	18.20%	15.00%	29.90%	47.40%	14.30%	19.10%
	4	14.30%	12.70%	25.70%	39.70%	76.20%	17.20%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Cascaded B	0	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	1	51.00%	50.90%	6.90%	3.80%	4.80%	42.50%
	2	35.90%	37.70%	52.50%	25.60%	16.70%	38.10%
	3	11.60%	11.10%	36.40%	65.40%	40.50%	17.20%
	4	1.60%	0.30%	4.20%	5.10%	38.10%	2.20%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Face ++ denotes the Faceplusplus method.
The **Percentage** value in each cell denotes the percentage of correctly estimated crowd size levels.
Darker cells denote higher accuracy than lighter ones.

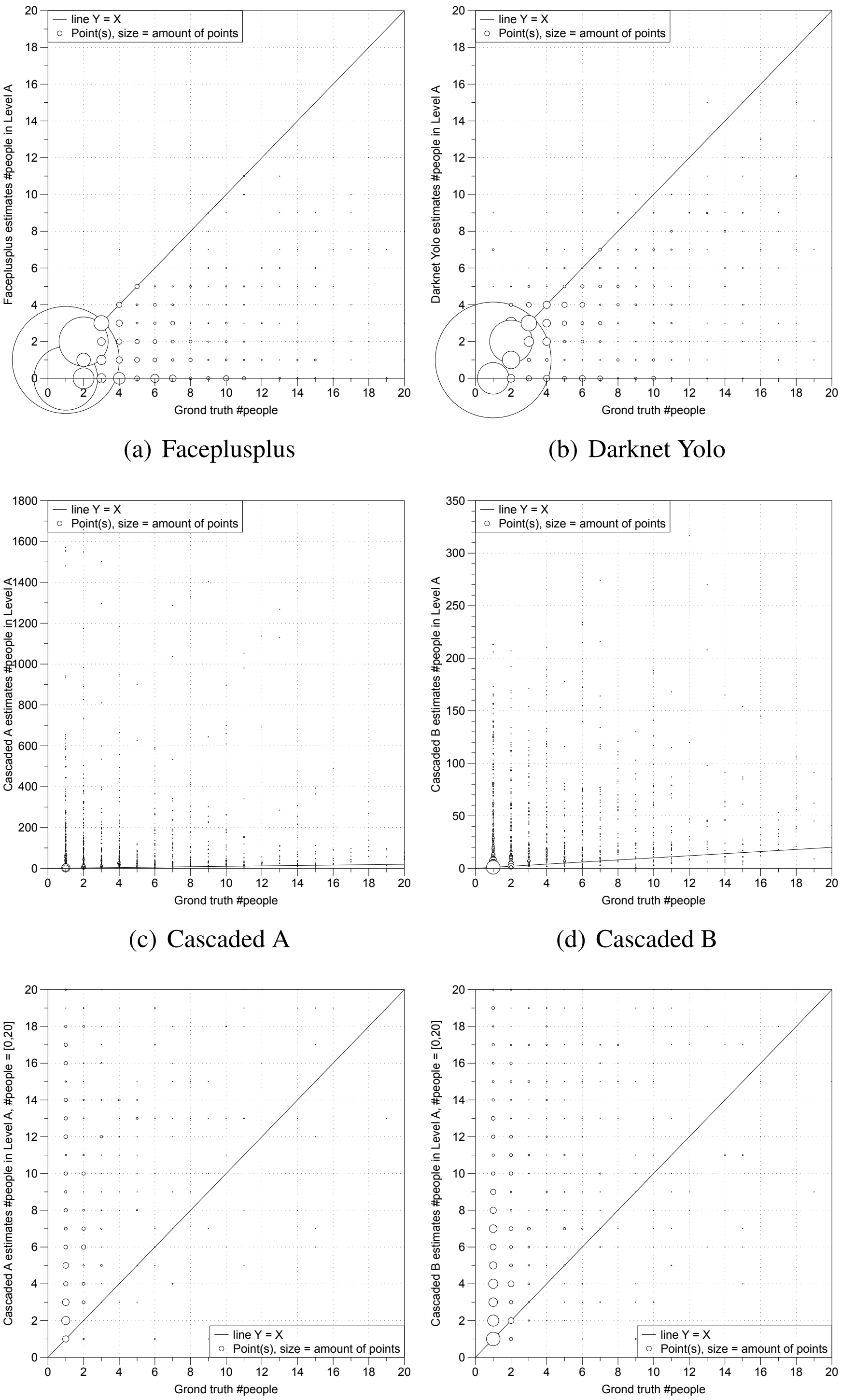
detected for counting people. In the meantime, as the constructed dataset contains more low dense images collected from social media, the estimation accuracy for Cascaded methods is obviously lower than Faceplusplus and Darknet Yolo.

Specific number of people in crowd size level 1 estimation

According to Table 4.6, Darknet Yolo reaches the highest estimation accuracy (38.09%) in the estimation of the specific number of people in crowd size level 1, closely followed by Faceplusplus (37.21%). The Cascaded A and B methods reach very low accuracy (2.05%, 4.88%). Similar to the observations in crowd size level estimation, the tendency of under- and over-estimation of different methods may caused by the different features they used for detecting people, as we described in the previous section.

To explore the relationship between the ground truth, the estimation value and amount of such estimation for each method, we plotted in Figure 4.6 with a ground truth value on the X axis, an estimated value on the Y axis and the number of corresponding estimation points in size. Points on the diagonal ($Y = X$) denote correct estimation. According to Figure 4.6(a) and Figure 4.6(b), the Faceplusplus and Darknet Yolo methods reach the highest accuracy in the range of 0 to 4. Instead, the accurate estimation for the Cascaded methods, according to Figure 4.6 (c,d,e,f), are distributed more equally than Faceplusplus and Darknet Yolo. This is consistent with the mechanism of different methods, i.e. Faceplusplus and Darknet Yolo are more feasible in low dense environment while Cascaded A and B are more feasible in high dense environment.

When comparing the two Cascaded methods, the accurate estimation in Cascaded B are more equally distributed than the Cascaded A. This is nature that most of social media image sent during city events are captured in outdoor event area, which are more feasible for the method Cascaded B, which are trained with busy street area in city events, than Cascaded A, which are trained with random pictures of city events.



(a) Faceplusplus
(b) Darknet Yolo
(c) Cascaded A
(d) Cascaded B
(e) Cascaded A, #people ¡= 20
(f) Cascaded B, #people ¡= 20

Figure 4.6: The distribution of specific number of people estimated in crowd size Level 1 for each method

4.8.2 Impact of image characteristics on crowd level estimation

We perform the second experiment to investigate the impact of image characteristics on crowd size level estimation for separate methods. The result is listed in Table 4.8 and 4.9, where each dark cell indicates an image characteristic (e.g. 'Condition') with current value (e.g. 'Outdoor') in the column has maximum positive impact for the corresponding method in the row. Namely, images with characteristics in such value have larger possibility to the correctly estimated the crowd size level by the corresponding method. For instance, the cell with impact scored 0.88 in column 'Parallel' indicates that images captured in 'view' (i.e. an image characteristic) of 'parallel' (i.e. the value of the image characteristic 'View') shows most positive impact on crowd size level estimation than any other value in view characteristic using the method Faceplusplus. Simply, it is more likely that Faceplusplus estimates the crowd size level correctly for images with taken in parallel view.

According to Table 4.8 and 4.9, the image characteristic 'people present' shows a negative impact for Faceplusplus and Darknet Yolo, but maximum positive impact for the Cascaded methods. This may be caused by the Faceplusplus and Darknet Yolo methods which tend to underestimate crowd size, and thus increased the correct estimation accuracy in processing images containing less people, in particular, no people.

We also found that indoor pictures show positive impact to all methods. It may be caused by indoor images containing less people which reduces the difficulties in crowd size estimation.

We observed that pictures taken in parallel view show higher positive impact for Faceplusplus and Darknet Yolo, while top view pictures are better interpreted by Cascaded methods. This may be caused that Faceplusplus and Darknet Yolo, counting people by faces and shapes, require more detail information about people than Cascaded methods, counting people through learned features as introduced in the previous chapter.

With regard to 'gathering distribution', all methods except Cascaded A shows higher estimation accuracy with fixed distribution of gatherings. This is natural that, compared with gradient distribution, the gatherings in fixed distribution contains less people and the people have a similar size in the image, which reduces the difficulties in detecting and counting people.

The findings show that all methods except Faceplusplus tend to correctly estimate the crowd size of images with gatherings. The Faceplusplus instead reaches a higher estimation accuracy with images containing clearest gatherings (so in level A) than no gatherings. It is natural that for other three methods, the gatherings which are in small size in the backend of the images increase the difficulties for crowd size estimation. However, as a face recognition based method, the Faceplusplus still can recognise small but clear faces (so in Level A) in the gatherings.

To assess the effectiveness of the impact of image characteristics on the crowd size estimation accuracy for each method, we tested the cross-validation performance of the by-product classifier constructed with impact of image characteristics. According to Table 4.9, both Faceplusplus and Darknet Yolo reach *F1_Score* at 0.86, while Cascaded A and B reach 0.76 and 0.74, respectively. It indicates that Faceplusplus and Darknet Yolo reach higher possibility (confidence) to produce correct crowd size level estimation than Cascaded methods when characteristics of images are in most positive impact values.

Table 4.8: The impact of image characteristics on crowd level estimation for each method, part 1

Method	Global image characteristics									Frontend image characteristics					
	People present		Condition			View				Selfie face					
	Yes	No	Outdoor	Indoor	Unknown	Parallel	Between	Top	Unknown	No selfie	Full face	Part face	Back face	Mixed face	Only body
Face ++	-0.719	0.719	0.526	0.689	-1.215	0.880	0.541	-0.206	-1.215	-0.279	1.197	0.114	-0.873	0.630	-0.787
Darknet Yolo	-0.366	0.366	0.457	0.758	-1.215	0.874	0.358	-0.017	-1.215	-0.442	0.693	0.329	0.100	0.157	-0.837
Cascaded A	1.354	-1.354	-0.139	0.398	-0.259	-0.098	0.167	0.190	-0.259	-0.096	0.034	0.083	0.259	-0.262	-0.018
Cascaded B	1.429	-1.429	-0.217	0.504	-0.287	-0.028	0.081	0.234	-0.287	-0.306	0.005	0.060	-0.142	-0.025	0.409

The **Face ++** denotes the Faceplusplus method.
Each **grey** cell indicates an image characteristic (e.g. 'Condition') with current value (e.g. 'Outdoor') in the column has maximum positive impact for the corresponding method in the row. Namely, images with characteristics in such value have larger possibility to the correctly estimated the crowd size level by the corresponding method.

Table 4.9: The impact of image characteristics on crowd level estimation for each method, part 2

Method	Backend image characteristics									Classification model performance
	Gathering distribution			Gathering clarity level						
	No gatherings	Fixed	Gradient	No gatherings	A	B	C	D	E	E(score)
Face ++	0.137	0.239	-0.375	0.288	0.376	0.080	-0.210	-0.111	-0.424	0.866
Darknet Yolo	0.020	0.329	-0.348	0.612	0.156	0.136	-0.235	-0.203	-0.466	0.869
Cascaded A	-0.323	0.005	0.318	0.479	0.034	0.068	-0.214	-0.241	-0.125	0.765
Cascaded B	-0.276	0.142	0.134	0.390	0.223	-0.066	-0.131	-0.239	-0.176	0.740

The **Face ++** denotes the Faceplusplus method.
The **E(score)** in the column of "Classification model performance" (the last column) denotes the average F1-score calculated in multi-folds validation in training of classification model.

4.9 Discussion

In this section, we discuss the research and findings in terms of the dataset, the effectiveness of algorithms, and how the social media data and the selected crowd size estimation methods fulfil the aim, i.e. estimating crowd size in city events.

The constructed dataset shows that social media data is indeed broadly available in city event area in large size, rather than data from sensors (e.g. Wi-Fi, counting system) which require extra resources in data collection. According to the constructed social media dataset in city events, 70% of images contain people. Thus, these images are valuable for crowd size estimation. Particularly, 20% of total images contain more than 20 people in each picture, which are essential in crowd size estimation for crowd management to reduce risks and incidents. However, the unbalance of number of images in terms of the way people are showing in the images (so as the different perspective the people are captured from camera, the selfie or panoramic pictures), and dense levels may affect the accuracy of methods in crowd size estimation, because methods identify and count people depending on such information.

With regard to social media data and the selected methods listed in Table 4.2, the findings show that direct methods (Faceplusplus and Darknet Yolo) tend to underestimate the crowd size while indirect methods (Cascaded A and B) tend to overestimate the crowd size. Also, direct methods are more effective with social media images in parallel view with full face and fixed clarity gatherings. Whereas, indirect methods are more effective when images are taken in top view with gatherings in gradient distribution. This may be explained by the mechanism of different types of methods, i.e. the direct methods detect and count people in an image by information of face or shape of people, are more feasible with images captured in parallel view (i.e. a viewpoint at more or less the same height as that of the people in the photo) with full faces of people and clearly observable gatherings around the people. In contrast, the indirect methods, such as Cascaded methods, which detect and count people through non-handcrafted features do not require the same information (e.g. face or shape of people) for crowd size estimation.

Moreover, the feasibility of different types of methods with social media images in different dense, i.e. direct methods fit low dense images while indirect methods fit high dense images, may indicate the crowd size estimation accuracy may be improved by selecting different methods under conditions which methods work best. For instance, to improve the crowd size estimation accuracy, we may classify the low dense and high dense images, or split frontend layer containing selfie people with backend layer containing gatherings in one image, and then assign low dense frontend images (layers) to direct methods and assign high dense backend images (layers) to indirect methods. To reach this, it is required to detect the image characteristics automatically and assign images with different characteristics to the different methods. Detecting the image characteristics automatically can be done with machine learning classification algorithms (Amerini et al., 2019; Wang et al., 2012b; McAuley & Leskovec, 2012), i.e. a classification model can be created and trained using annotated images, and the model can further be used for categorising images according to different characteristics, such as taken indoor or outdoor, whether people are present, camera viewpoint, whether it is a selfie picture, gathering distribution and clarity level.

4.10 Summary and Conclusion

Knowing the crowd size is essential for crowd safety when managing the crowd. Conventional solutions to derive such information depend on manual observations, which are expensive, prone to observation biases, and not suitable for global observations.

In this paper, we investigate the accuracy of four methods to estimate crowd size using social media images in city events, and we also investigate the impact of image characteristics on the estimation effectiveness for different methods.

To perform this research, we select four methods for crowd size estimation analysis from two different types. We created a dataset consisting of social media images collected from various events and major activities. Each image is annotated with a set of image characteristics and crowd size as ground truth. This dataset has been used for investigating the crowd size estimation accuracy of selected methods and the impact of image characteristics on the crowd size estimation.

Findings show that direct methods (Faceplusplus and Darknet Yolo) reach better estimation accuracy than indirect methods (Cascaded method A and B). Specifically, Darknet Yolo reaches the highest accuracy in crowd size level estimation (72.01%) and in estimating the specific number of people when less than 20 (38.09%). The findings indicate that, social media images taken in parallel view with selfie people in full face and gatherings in fixed distributed are more feasible to Faceplusplus and Darknet Yolo for crowd size estimation, while images taken from top view with gatherings in gradient distribution are more suitable to Cascaded methods. We recommend to use Darknet Yolo method to estimate and predict the crowd size in city events based on social media images, in terms of levels of crowd size as well as specific number of people if it is a low dense environment.

Results of this research may be influenced by the construction bias of the social media dataset, which are introduced by the diverse characteristics of city events. Though we considered a set of event characteristics in dataset construction, it may not be able to cover all diversities in the reality. Thus, the crowd size estimation performance using social media images may be affected in city events beyond the scope of consideration in terms of diversity of event characteristics. Moreover, the bias in social media usage in terms of users' age and gender may also affect crowd size estimation using social media. For instance, knowing that social media is more popular in younger generation (Yang et al., 2016; Gong et al., 2018a), city events with less younger participants may generate less social media images, which may not sufficient for training and improving the crowd size estimation methods, and also not sufficient for methods to estimate the crowd size during events. Therefore, it may affect crowd size estimation for crowd management.

In future work, we plan to propose and explore more methods for crowd size estimation using social media images, such as hybrid methods that integrate the advances of direct and indirect methods, in such a way to improve the estimation effectiveness for crowd management. Also, we intend to enlarge the social media dataset by adding more diverse events and activities, and by adding more annotated image characteristics. Last but not least, we will investigate the feasible approach to derive image characteristics automatically from social media dataset.

Chapter 5

Using Social Media for Attendees Density Estimation in City-Scale Events

In the previous chapter, we explored methods for estimating the number of people (crowd size) in the crowds using social media images collected during city events. In this chapter, we estimate the density of attendees using social media data. This answers the fourth research question, i.e. **RQ4. To what extent are social media data able to estimate the density of people in city events?**

To this end, we propose three classes of density estimation strategies (i.e. geo-based, speed-based and flow-based), inspired by elements of pedestrian traffic flow theory that were successfully assessed during city-scale events. We study the performance of these strategies in the context of SAIL Amsterdam 2015 (Sail) and Kingsday Amsterdam 2016 (Kingsday), two city-scale events that respectively attracted 2 and 1.5 million attendees in the span of 5 days and 1 day, respectively. We defined four experimental areas for the Sail event and one for the Kingsday event, and compare density estimates from social media data with measures obtained from counting systems and Wi-Fi sensors.

Results show the potential of solutions using elements from pedestrian traffic flow theory, which yielded estimates with strong temporal correlations with the sensor observation, and limited average errors.

This chapter is published as a journal article: Gong, V. X., Yang, J., Daamen, W., Bozzon, A., Hoogendoorn, S., & Houben, G. J. (2018). Using social media for attendees density estimation in city-scale events. *IEEE Access*, 6, 36325-36340.

5.1 Introduction

As cities battle for global importance and influence, city-scale public events are becoming an important weapon of choice to foster tourism and economic growth. Olympic games, thematic exhibitions, and national celebrations are examples of city-scale events that take place in vast urban areas, and attract large amounts of attendees within short time spans. The scale and intensity of these happenings demand for technological solutions able to support relevant stakeholders (e.g. event organisers, public and safety authorities, attendees) with the monitoring of an event's state with respect to the crowd.

For instance, it is common for public authorities to monitor the number of attendees present in a given event terrain, to promptly identify capacity issues and minimise the risk of incidents due to overcrowding – stampedes are more likely to occur in high-density crowds (Helbing et al., 2005). The estimation of attendees *density* requires a measurement infrastructure that is characterised by stringent requirements in terms of spatial resolution, temporal resolution, and accuracy. These measurement activities are typically performed by personnel operating on the event terrain (Wirz et al., 2012); the data they provide is however temporally scarce, spatially non-uniform, and often subjective.

Ad-hoc sensing infrastructures – such as counting system and Wi-Fi sensors – or pre-existing communication infrastructures – such as mobile phone networks – are an automatic solution for the real-time measurement of the number of individuals and/or connected devices present in a given area (Daamen et al., 2016). Their widespread adoption is however constrained by economical and operational limitations. Counting system infrastructures are expensive to set-up and operate; their monitoring capability is limited to a fixed and relatively small area; as counting is performed by means of computer vision algorithms trained to recognize human faces, heads or shoulders, their accuracy decreases in non-standard operational conditions – for instance, when it becomes too crowded, or when adverse meteorological conditions force people to use umbrellas. The accuracy of Wi-Fi sensors is clearly dependent on issues such as technological penetration, technology of devices; and data from mobile communication may be only available at coarse-grained resolution due to privacy or technological limitations.

Social media data produced by platforms like Twitter or Instagram are increasingly used to study urban-related problems (Cranshaw et al., 2012; Hasan et al., 2013; Quercia et al., 2015a), and to monitor the on-line liveness of city-scale events (Lee & Sumiya, 2010; Balduini et al., 2013). Their popularity is certainly due to their availability, ease of access, real-timeliness, and geographical annotation. On the other hand, social media data suffer from known limitations in terms of representativeness of the targeted population, and (spatial and temporal) sparsity. Intuitively, not all attendees feel compelled to share their experience on social media, or are active on such platforms; also, event areas are differently attractive; and the event is not equally engaging over time.

As a result, there is a lack of scientific knowledge about the suitability of social media as a data source for density estimation. In this paper, we aim at filling this knowledge gap by studying how micro-posts harvested from social media can be used during city-scale events to estimate the density of attendees stationing in – or moving through – a given terrain. We formalise the problem in a probabilistic framework, and calculate the likelihood of event attendees to be present in the targeted event terrain within a given time span. Inspired by methods of pedestrian traffic flow theory successfully tested in crowd monitoring applications (Yuan et al., 2016), we propose 3 density estimation strategies: *geo-based*, *speed-based*, and *flow-based* strategy.

The assessment of the performance of these strategies in real-world settings is a challenge per-se, and it is often neglected in existing studies. This work contributes the results of an analysis performed on two large-scale sensing infrastructures, that we set-up in the city of Amsterdam during SAIL 2015 (Sail) - the largest free nautical event in the world, and King's Day 2016 (Kingsday) - the national King's birthday event, held once a year, and attracting millions of people. During the Sail event, we focused on 4 terrains located along a walking route close to where most tall ships were moored; during King's day, we focused on 1 terrain in the south of Amsterdam, covering a busy square between Station Amsterdam Zuid and World Trade Center (WTC) with various shops and restaurants around. These 5 event terrains are characterised by different morphology and relevance to the activities of both events.

We then compared the density values estimated from social media data with the measures obtained from the sensing infrastructure. Results show that the proposed density estimation strategies are able to cope with data sparsity issues typical of geo-referenced social media. Errors in density estimation are in the range of 1-2 order of magnitudes, but with strong temporal correlations with measures obtained from the sensing infrastructure. Finally, we show that density estimation is influenced by the characteristics (e.g.

morphological and functional) and the traffic status of the monitored terrain. We stress the importance of a systematic comparison with real-world data, and the challenging nature of our experimental setting: in our work we are able to provide novel insights into the suitability of social media as a data source for density estimation, and to ground them against measurements from state-of-the art pedestrian traffic flow measurement infrastructures.

The remainder of this work is organised as follows: in the next section, related works are discussed. In Section 5.3, we propose our method to tackle this problem, followed by experimental setup for two cases in Section 5.4. The results of experiments are presents in Section 5.5 and discussed in Section 5.6. The conclusions including future research of this article is in Section 5.7.

5.2 Related Work

A growing number of studies investigates pedestrian behaviour models aiming at developing systems to automatically identify overcrowding during city-scale events. Wirz et al. (2012) propose a pedestrian-behaviour model to infer crowd conditions in city-scale events based on GPS location traces. Blanke et al. (2014) study crowd mobility dynamics in city-scale events using GPS data. Weppner & Lukowicz (2013) study the problem of density estimation by Bluetooth scans with mobile phones. However, fewer works attempt to make use of social media data to provide insights into attendees' behaviour during city-scale events, while numerous recent works (Cheng et al., 2011; Cranshaw et al., 2012; Quercia et al., 2015b,a) provide evidence that social media data can give semantically rich insights into the spatio-temporal dynamics of urban areas.

Botta et al. (2015) show evidence of a relationship between the number of attendees at a given location at a given time with their social activities. They performed a correlation analysis of the number of attendees in two cases, a football stadium and an airport, with regard to their social media usage on Twitter, mobile calls and SMS activities on 11 event days in a city. It showed that data generated through interaction between people can be used to extrapolate the number of people in a given location at a given time, which may be valuable for business and policy makers. However, the purpose of their work is slightly different from ours. In our work we also use social media as data source to estimate the number of attendees at a location during a given time period. In order to provide valuable information for crowd management, we target on a more fine-grained analysis, i.e. in a hourly basis and within several specific terrains. This also leads us to deal with social media sparsity during a short time and within a small space. Besides, we also looking into insights from social media data to interpret the estimation result.

Liang et al. (2013) establish a model to calculate the volume of event attendees through social media, considering the number of check-in users and the duration of their stay in an event. Their model uses check-in and check-out number of social media users to estimate population. The check-in number of people is calculated by the number of posts sent from a location. While the check-out number of people is calculated through the number of check-in people with the length of duration each people stay in the event. The duration time is estimated using timestamps between multiple posts sent by one user. The advantage of this model is that it transfers a population modelling problem into a temporal duration estimation problem making use of timestamp information of multiple posts sent by one user. Similar to our method, to tackle the social media sparsity authors make use of the duration information to estimate an emission rate, i.e. a probability of a person sending a post during an event in a crowd. However, using the duration information as signal for estimation population of a crowd will introduce bias as fewer people sent multiple posts in one day, which reduces the precision of the estimation. To avoid this risk, in our method, instead of using the duration information, we construct the probability by loosing the temporal and spatial limitation to count people nearby.

Georgiev et al. (2014) further investigate factors which influence people participating in an event using social media data. It shows evidence that friends' co-attendance and the popularity of the event are dominating factors. In our work, we further interpret results using profile information derived from social media data, such as age, gender, city-role, and PoI preference of users.

5.3 Estimating Attendees Density from Social Media Data

This section introduces the problem of attendees density estimation, and presents our proposed solutions. First, we introduce concepts from pedestrian traffic flow theory useful in the context of density estimation. Then, we describe three classes of density estimation strategies, namely: 1) *geo-based* strategies, operating only on social media data; 2) *speed-based* strategies, which estimate density by considering the travel speed (i.e. distance covered per unit of time) of attendees on the event terrain; and 3) *flow-based* strategies, that consider travel flow information (i.e. number of attendees passing a reference point per unit of time).

5.3.1 Pedestrian Traffic State Variables

In pedestrian traffic flow theory (May, 1990; Daganzo, 1997; Daamen et al., 2005), one of the fundamental characteristics of a moving population, from a macroscopic point of view, is the average flow $q = vk$. Given the average walking speed v (m/s) and the average density k (P/m^2), the flow q (P/ms) is defined as their product.

Density is a property related to a terrain where the event takes place, i.e. a shaped space formed with boundaries defined by a set of coordinates. To simplify the discussion, we assume event terrains to have rectangular shapes as in Figure 5.1. Consider an event terrain e having area A_e. The density is defined as the number of attendees P per unit area of the event terrain at a certain moment in time ts, and is formalized as follows (Duives et al., 2015):

$$k(e,t_s) = \frac{\mathcal{P}(t_s)}{A_e} \tag{5.1}$$

$\mathcal{P}(t_s)$ denotes the number of attendees at the terrain e at t_s.

Speed is the distance of attendees' movement per unit time. Consider an attendee crossing a whole terrain e during the time window $[t_1,t_2)$, the speed is formally defined as:

$$v(e,t_1,t_2) = \frac{L_e}{|[t_1,t_2)|} \tag{5.2}$$

where L_e is the distance covered by the attendee when moving through the terrain e. When considering multiple attendees moving through a terrain in different time windows, we could obtain a distribution of speed as a property associated to the terrain, denoted as $\mathbb{V}(e)$.

For an event terrain e, the net flow of attendees traversing a terrain boundary b_e during the time window $[t_1,t_2)$ is defined as:

$$q(t_1,t_2) = \frac{\mathcal{P}_{in}(t_1,t_2) - \mathcal{P}_{out}(t_1,t_2)}{|[t_1,t_2)|} \tag{5.3}$$

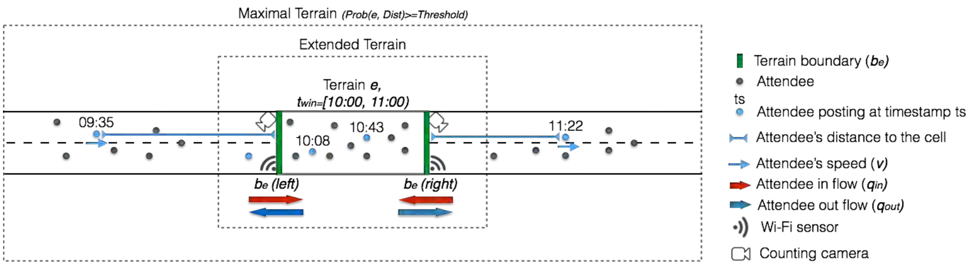


Figure 5.1: Illustration of geo-, speed- and flow-based density estimation methods. To estimate attendees density in the terrain during the time window $t_{win} = [10{:}00, 11{:}00)$, geo-based density estimation method considers the number of users posting at least once within the terrain ($\widehat{k_1}$) or within the extended terrain ($\widehat{k_2}$) during t_{win}; the speed-based density estimation method ($\widehat{k_3}$) considers attendees travel speed to account for attendees that could potentially be present in the terrain during t_{win}, but that post on social media in a location within walking distance to the terrain; the flow-based density estimation method ($\widehat{k_4}$) further considers attendees flow information produced by the sensing infrastructure.

$\mathcal{P}_{in}(t_1,t_2)$ and $\mathcal{P}_{out}(t_1,t_2)$ denote, from t_1 to t_2, the number of attendees moving into the terrain through this boundary, and the number of attendees moving out the terrain, respectively. A flow value $q(t_1,t_2) > 0$ indicates that through b_e the number of attendees entering the terrain exceed the attendees that exit it from t_1 to t_2; otherwise, $q(t_1,t_2) < 0$.

5.3.2 Geo-based Density Estimation

Density, as defined in Eq. 5.1, can be measured using traditional sensing infrastructures (e.g. counting systems and Wi-Fi sensors) by means of state-of-the-art methods (Yuan et al., 2016).

The sparse nature of social media data, however, calls for different ways to measure density. Intuitively, given an arbitrary event terrain e (e.g. a square, a venue), the number of people performing social media activity at a give time instant t_s is normally rather small. To account for such sparsity, we modify the definition of density by considering it a property associated to a *time span* $t_{win} = [t_{start}, t_{end})$. We therefore formalise density measured through social media data as follows:

$$\widehat{k_1}(e, t_{win}) = \frac{|\{u|\forall u \in U, p_u(t_{win}) \geq 1\}|}{A_e} \tag{5.4}$$

where U is the set of event attendees generating social media activities at the location of the event terrain and $p_u(t_{win})$ denotes the number of posts the social media user u post in t_{win}. The density $\widehat{k_1}$ of a terrain e in the time window t_{win} is therefore calculated as the number of users posting *at least one* micro-post in the targeted area during the considered time window. Considering sparsity of geo-referenced social media data, we choose a time window of one hour. Figure 5.1 shows an example estimating the density of the terrain for time window $t_{win} = [10{:}00, 11{:}00)$ considering social media sparse. We leave the investigation of density estimation in shorter time windows to future work.

While increasing temporal boundaries for density calculation, the previous definition puts a very strict constraint on the geographical boundary of the terrain of interest. Attendees could perform social media activity in close proximity to the terrain area. Their communication device could also introduce localization errors due to technical[1] or environmental (e.g. signal blockage, proximity to tall buildings) issues. These errors can range from dozens of meters[2] to even more than 100 meters.[3] [4]

To account for such uncertainty, we consider a second definition of density where the boundaries of the considered terrain area are extended by 111.32 meters[5] in each direction. The resulting density measurement is expressed as:

$$\widehat{k_2}(e, t_{win}) = \frac{\{u|\forall u \in U, p_u(t_{win}) \geq 1\}}{A_e^{extend}} \tag{5.5}$$

5.3.3 Speed-based Density Estimation

Though the second definition in the previous section accounts for attendees sent posts in the terrain e or in the extended terrain e during the time span of interest, it does not account attendees who could have been active before

[1] https://tnp.uservoice.com/knowledgebase/articles/1117027-gps-location-errors
[2] https://www.gps.gov/systems/gps/performance/accuracy/
[3] http://www.radio-electronics.com/info/satellite/gps/accuracy-errors-precision.php
[4] https://msu.edu/~brook/publications/prec_ag/oct1998.htm
[5] 111.32 meters are equivalent to a decimal degree precision of 3 decimal places: https://en.wikipedia.org/wiki/Decimal_degrees

entering e, or after leaving it. By considering attendees travel speed, it is possible to account for people that could potentially be present in e in the time span of interest, but posted on social media in a location within walking distance.

Pedestrian speed is known to approximately follow a normal Gaussian distribution (Chandra & Bharti, 2013). City-scale events can be very crowded: with lots of activities taking place on the event terrains, the motion of pedestrian can be relatively slow. This is the experimental conditions in the "Precinct" scenario of Chandra & Bharti (2013) where $\mathbb{V}(e) \sim N(0.97, 0.21^2)$. We therefore use this result as the assumed pedestrian speed distribution in our study. We leave the robust analysis with respect to the assumption of parameters as well as the assumption in different terrains as future work. We include a parameter Δt that constrains the temporal scope of our model: only users whose posts are detected in the time span $[t_{start} - \Delta t, t_{end} + \Delta t)$ (where $t_{win} = [t_{start}, t_{end}))$ are to be considered. As an example, for the terrain in Figure 5.1 and the time window $t_{win} = [10{:}00, 11{:}00)$, we consider an extended time span $[09{:}30, 11{:}30)$ (i.e. $\Delta t = 30$ minutes) to account for attendees' travel speed. Attendees posting during this time span, e.g. posting at 09:35, could be present in the terrain during $[10{:}00, 11{:}00)$, are therefore included in the density estimation.

Given the speed distribution and the scoped amount of time, attendees active on social media outside the terrain e before t_{start} (respectively, after t_{end}) will have a probability of being in e within t_{win} that is related to their distance.

Assume a user u to be active at a distance d from the event terrain of interest. We use pdf to denote the probability density function of traveling speed. Intuitively speaking, the user should have a speed of at least $\frac{d}{\Delta t}$ in order to reach the terrain e within Δt. Therefore the probability equals to the probability of $v = \frac{d}{\Delta t}$ in the inverse cumulative distribution function of speed distribution. This means that a social media user is more likely to reach the terrain within a certain time window when performing an activity with small distance from considered terrain. The probability of being in the terrain within Δt can be calculated as:

$$\begin{aligned} P_{\Delta t}(e,d) &= P(v(e) \geq \frac{d}{\Delta t}) \\ &= \int_v pdf(v(e) \geq \frac{d}{\Delta t}) \end{aligned} \tag{5.6}$$

Assuming that at the same location with distance d to the terrain there are $N(d)$ attendees active on social media, then $N(d) \times P_{\Delta t}(e,d)$ of them will possibly be in the terrain during t_{win}. When considering users at locations with different distances from the terrain, the number of users that could contribute to the density of the terrain in the considered time span can be calculated as:

$$\begin{aligned}\widehat{k_3}(e,t_{win}) =& \frac{1}{A_e}\Bigg\{\{u|\forall u \in U, p_u(t_{win}) \geq 1\} \\ &+ \int_d N(d) \times P_{\Delta t}(e,d)\Bigg\}\end{aligned} \tag{5.7}$$

5.3.4 Flow-based Density Estimation

Data about attendee flows (i.e. number of attendees traversing the boundaries of a terrain per unit of time) could also be used to support attendees' density estimation. Such flow information can be obtained by counting systems and/or Wi-Fi sensors, as illustrated in Figure 5.1. Values of $q(b_e,t_1,t_2)$ for other moments of time, such as the previous day, previous week, or during the event on the same day last edition, could be used to scale up attendees' density in the terrain by scaling the probability $P(e,d)$ in Eq. 5.6 before t_{start} (or after t_{end}) according to previous traffic conditions. To model this, we consider for each terrain boundary b_e the number of attendees 1) active before t_{start} ($[t_{start} - \Delta t, t_{start})$) and 2) after t_{end} ($[t_{end}, t_{end} + \Delta t)$). We use $c_{bf}(b_e)$ and $c_{af}(b_e)$ to denote the scaling factors for boundary b_e considering attendees active before t_{start} and after t_{end}, respectively. In addition, $N_{bf}(d)$ and $N_{af}(d)$ denote the number of social media users with distance d to the terrain before t_{start} and after t_{end}. The estimated density is calculated as follows:

$$\begin{aligned}\widehat{k_4}(e,t_{win}) =& \frac{1}{A_e}\Bigg\{\{u|\forall u \in U, p_u(t_{win}) \geq 1\}+ \\ &\sum_e \Bigg(c_{bf}(b_e)\int_d N_{bf}(d) \times P_{\Delta t}(e,d)+ \\ &c_{af}(b_e)\int_d N_{af}(d) \times P_{\Delta t}(e,d)\Bigg)\Bigg\}\end{aligned} \tag{5.8}$$

The scaling factor $c_{bf}(b_e)$ and $c_{af}(b_e)$ for each boundary b_e are calculated as in Eq. 5.9, to respectively account for activities performed before or after the considered time span. In the equation, $t_s = t_{start}$ and $t_e = t_{end}$.

$$
\begin{aligned}
c_{bf}(b_e) &= \begin{cases} \frac{\mathcal{F}(b_e, t_s - \Delta, t_s)}{\int_d N_{bf}(d) \times P_{\Delta t}(e,d)}, & \text{if } \mathcal{F}(b_e, t_s - \Delta t, t_s) > 0 \\ 0, & \text{otherwise} \end{cases} \\
c_{af}(b_e) &= \begin{cases} \frac{|\mathcal{F}(b_e, t_e, t_e + \Delta t)|}{\int_d N_{af}(d) \times P_{\Delta t}(e,d)}, & \text{if } \mathcal{F}(b_e, t_e, t_e + \Delta t) < 0 \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
\tag{5.9}
$$

Let us first consider the case of attendees active outside the terrain during $[t_{start} - \Delta t, t_{start})$. The scaling factor $c_{bf}(b_e)$ assumes a positive value when $\mathcal{F}(b_e, t_{start} - \Delta t, t_{start}) > 0$, i.e. when, in the considered time period there are more attendees entering the terrain than leaving it. When, on the other hand, $\mathcal{F}(b_e, t_{start} - \Delta t, t_{start}) < 0$, i.e. there are more attendees leaving the terrain than entering it, their impact can be modelled as $c_{bf}(b_e) = 0$, that is, no additional attendees active on social media should be counted in estimating the density of the terrain during $[t_{start}, t_{end})$.

When attendees are active outside the terrain during $[t_{end}, t_{end} + \Delta t)$ (i.e. after the considered time span), the positive and negative of scaling factor $c_{af}(b_e)$ are the other way around.

5.4 Experimental Setup

This section describes the experimental infrastructure designed and implemented in our work.

We performed our studies in the context of two events city-scale events, the SAIL Amsterdam 2015 nautical event (Sail) and Kingsday Amsterdam 2016 national holiday (Kingsday). First, we elaborate reasons for selecting these two events. Then, we provide a brief introduction of each event, and introduce their terrains focused upon in the experiment. Further, we detail the 4 experimental testing definitions. Finally, we introduce the sensor and social media data collection infrastructure, and the metrics used to compare the performance of our density estimation methods (working on social media data) against the density measurement performed through the sensing infrastructure, here interpreted as ground truth.

5.4.1 Event selection

The areas affected by Sail and Kingsday are shown in Figure 5.2. In the attempt of broadening the scope and validity of our work, we selected events sharing similar properties. Both Sail and Kingsday are 1) *city-scale* events

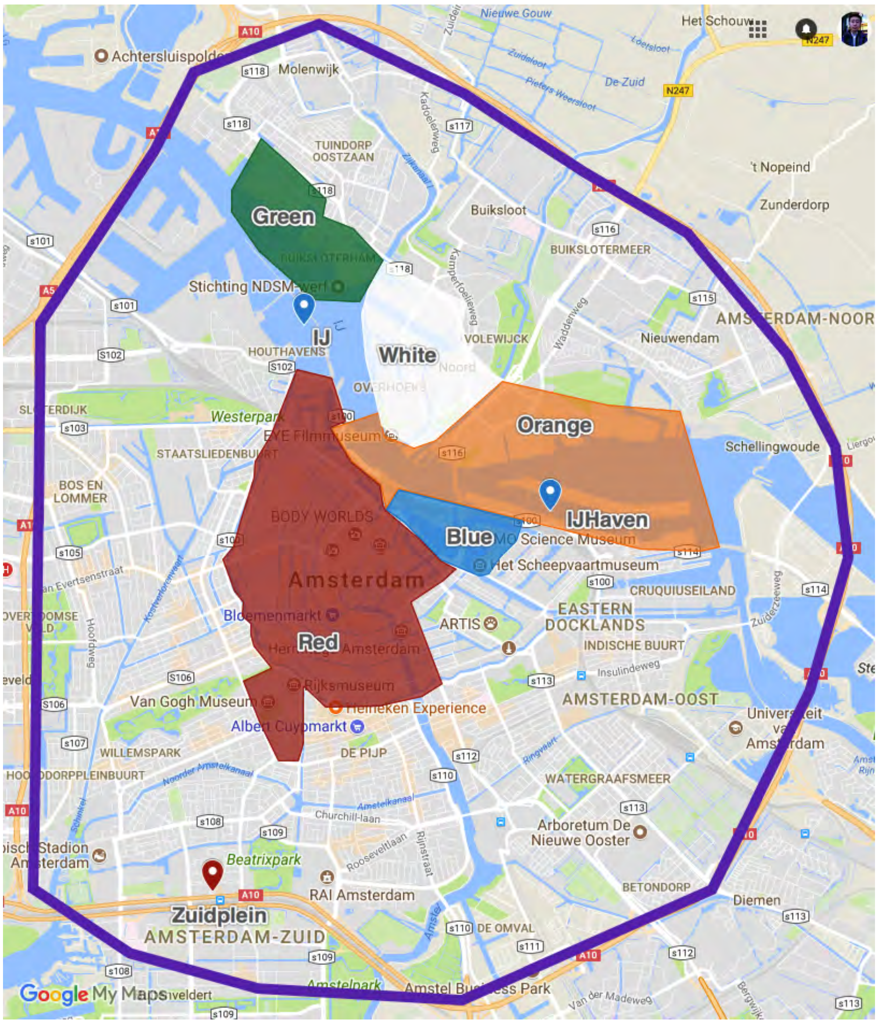
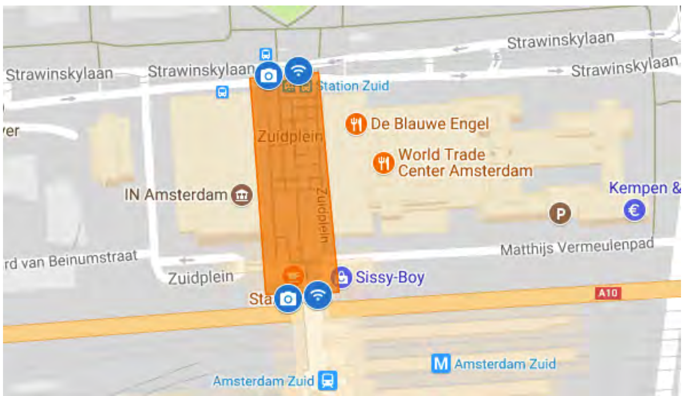


Figure 5.2: *Location of targeted terrains in Sail 2015 and Kingsday 2016 in Amsterdam. Most of activities during the Sail event took place in 5 colored oceans (areas), i.e. Orange, White, Blue, Green and Red Oceans. Activities during Kingsday took place in the whole city of Amsterdam (area bounded by dark blue line). Marked locations indicate where the terrains considered in the research are located. Terrains of the Sail event are located around the IJhaven (Blue marker), while the terrain on Kingsday is located at Zuidplein (Red marker).*

(a) Pictures from Sail event

(b) Terrains of Sail event

(c) Pictures from Kingsday event



(d) Terrain of Kingsday event

Figure 5.3: Sail Amsterdam 2015 and Kingsday Amsterdam 2016 selected for the experiment

taking place in the same urban environment; 2) *planned*, *temporally constrained*, and thoroughly organized (in contrast to seasonal events, such as Christmas shopping, or serendipitous events, like protests); 3) *popular*, as they are known to attract large crowds, regardless of weather conditions; and 4) *generalist*, and they attract diverse demographics. At the same time, the two events also have important differences, such as 1) *duration*, as Sail lasts for 5 days, ending in a weekend. While, Kingsday is a single-day event, and a public holiday, with celebrations starting from one day before the event day and last for day after it; 2) *topic*, being Sail a naval event (offering, for instance tall-ship exhibition, nautical history experience, fireworks show), while Kingsday is a recurrent national celebration, which offers a boat parade, free market and parties; 3) *event terrain*, with Sail activities centred around the IJhaven area (where ships docked), while Kingsday activities are scattered throughout the city.

More details about events and their terrains for this experiment are introduced in the following sub-sections.

The SAIL Amsterdam 2015 Nautical Event

SAIL [6] is the largest free nautical event in the world. It takes place every five years in the city of Amsterdam, being the largest public event in the Netherlands. It hosts tall ships from all over the world, moored in the eastern harbour of the city *IJHaven* (IJ harbour) and across the IJ river for attendees coming from all over the world to see and visit. The 2015 edition of SAIL took place from August 19 until August 23, and attracted in total more than 2 million attendees. A high-level view of the area of Amsterdam where the event took place is depicted in Figure 5.3(c).

The event organisers predefined several walking routes for the attendees to follow. A detailed map of the SAIL event, its routes, and its point of interest is available on the event website[7]. The routes included streets facing the ships' docking areas. Each street is characterised by different morphology (length and width of attendees routes), facilities (e.g. toilets, information desks) and exposure to the main attractions. The main route, called *Orange* route, started from the Amsterdam Centraal station (*Ruijterkade*); it then proceeded east towards the end of the IJHaven passing by the *Veemkade*; to continue north around the *Java Eiland*, first traversing the *Javakade*, and

[6]https://www.sail.nl/EN-2015
[7]https://www.sail.nl/media/644212/sail_perskaart_1400_990.pdf

then heading back through the *Sumatrakade*. The streets in proximity to the main attractions hosted stages (e.g. from sponsors) and markets. Buildings close to the event hosted concerts and other initiatives, and, in general, the part of the city nearby the IJHaven transformed to accommodate the event and its attendees.

The weather has been warm and dry for the whole duration of SAIL 2015. The programme included events spanning all five days. August 19 was mainly characterised by the *SAIL-in* parade: the first ships started at 10:00 in IJmuiden and arrived around 14:00 in Amsterdam, while the last ships entered Amsterdam around 17:00. All tall ships entered Amsterdam via the North See Canal, to then dock in the IJHaven. During the following three days, the tall ships were open for visits from 10:00 till 11:00. They then departed on August 23 during the closing *SAIL-out* events. Every day, a firework exhibition took place in the IJHaven around 11:00.

The authors were active in the crowd control room of SAIL 2015, and therefore could witness the evolution of the event. The fourth day (Saturday) was expected to be most crowded, mainly because of locals having their day off. Some crowd management measures have been applied, especially on Saturday afternoon. The *Veemkade*, where most of the tall ships were anchored, was very crowded, with queues forming to access the tall ships. Around stages and other points of interest, people stood still to enjoy music, to have social interactions with other attendees, or to consume food and drinks. Also, the *Javakade*, where people walk through narrow pedestrian bridge and watch tall ships docked in IJHaven, was very crowded.

We focused on four event terrains in Sail for this experiment, highlighted in Fig. 5.3(b):

- **Terrain 1: Ruijterkade** (Blue. Length: 657m. Width: 109m. Area: 6.12ha): the terrain is located at the north of the Amsterdam Centraal station. It continuously serves people using public transport services (the train station, or ferries directed to the northern part of Amsterdam). During SAIL, it served as a main access point to the event. The terrain hosted no relevant points of interest.

- **Terrain 2: Veemkade** (Turquoise. Length: 485m. Width: 71m. Area: 3.41ha): main terrain of the event, where most of the ships were docked. The area hosts offices, bars and restaurants, and some private residence. The terrain gave access to the majority of docked boats.

- **Terrain 3: Javakade** (Red. Length: 617m. Width: 78m. Area: 4.80ha): located on the Java Island, the street directly faces the IJHaven. The terrain is residential, with no recreational businesses. Small pedestrian bridges connect areas separated by canals. The terrain gave access to several docked boats.

- **Terrain 4: Sumatrakade** (Green. Length: 253m. Width: 56m. Area: 1.38ha): located on the Java Island, facing the IJ. The terrain hosted less attractions, compared to the previous two terrains and gave access to only few boats.

During the event, all locations were devoted to pedestrian and bicycles. Cyclist traffic was reduced during the more crowded hours.

The Kingsday Amsterdam 2016 Event

Kingsday is a national holiday held each year in April 26th in major cities in the Netherlands. It is the birthday of King Willem-Alexander, celebrated with joyful open air festivities. People join this yearly event with their families and friends. In 2016, the King's day celebration attracted more than 1.5 million people in Amsterdam, including Dutch tourists and an organic amount of foreign tourists.

Though a one day public holiday, Kingsday is certainly not a day of rest. The celebrations start on the eve of King's day - named as King's night. Parties, music, and carnival atmosphere continuing throughout the city till the end of the big day. Following the King's Night, the major activities taking place on King's day are free market, boat parade, and gay parties. On King's day morning from 6:00 onwards, the citywide street market in Amsterdam facilitates attendees into trading of their secondhand wars on the streets and in the parks, creating one of the world's largest flea markets. South Amsterdam has the biggest market. In the Jordaan, a crowded market is carried out with folk singers music. Markets in the Vondelpark are dedicated for kids to trade their toys or clothes. From 13:00 onwards, canals are packed with boat parties, with boats sailing along the canals throughout the city with great party vibrations on it. Various street parties and sub-events are carried out in the city with everyone wearing orange. Gay parties are held around Westermarkt and Reguliersdwarsstraat. Besides parties, several big museums are open for people who would like to experience the culture and history.

Kingsday activities occur in the whole city. Pedestrian areas nearby transportation hubs are particularly crowded as people were gathering there and enjoying various activities. We focused on one terrain shown in Fig.5.3(d).

- **Terrain: Zuidplein**: the terrain is the forecourt of the station Amsterdam Zuid. It is a popular pedestrian square located between Station of Amsterdam Zuid and the Strawinsky Avenue surrounded by the World Trade Center (WTC) in the south of Amsterdam. Around the square, there are various shops, sandwiches and other amenities, attracting lots of people. It is a major pedestrian terrain connecting Amsterdam OUD-Zuid, with the CBD area, and Station of Amsterdam Zuid. Nearby, there are two large events in the RAI and the Olympic stadium, which generates large pedestrian flows through this station.

5.4.2 Experimental Conditions

We investigate in this paper the properties and performance of the following density estimation methods:

- $\widehat{k_1}$: geo-based density estimation, considering the exact geographical boundaries of the targeted terrain;
- $\widehat{k_2}$: geo-based density estimation, considering the extended boundaries of the targeted terrain;
- $\widehat{k_3}$: speed-based density estimation, using the pedestrian speed distribution suggested by (Chandra & Bharti, 2013) to calculate the probability of social media activities to occur in the targeted terrain;
- $\widehat{k_4}$: flow-based density estimation, using flow estimated through the sensing infrastructure to scale the probability of social media activities.

All methods estimate density from social media data on an hourly basis.

5.4.3 Data Collection

Our experiment took place during the first four days of the SAIL event, and the whole day of the Kingsday event, focusing on the terrains introduced in the previous sections, i.e. the Ruijterkade, Veemkade, Javakade, Sumatrakade for the Sail event, and the Zuidplein for the Kingsday event.

Table 5.1: Sensing infrastructure and social media monitoring on targeted terrains

	Terrain	Counting systems	Wi-Fi	Social media	
				Twit.	Inst.
Sail	Ruijterkade	Single	Both	Y	Y
	Veemkade	Both	Both	Y	Y
	Javakade	Single	Both	Y	Y
	Sumatrakade	Single	Both	Y	Y
Kingsday	Zuidplein	Both	Both	Y	Y

Twit. = Twitter, Inst. = Instagram

Single: the sensor is equipped on one boundary of this terrain.

Both: the sensor is equipped on both two boundaries of this terrain.

Y: data of this social media network is collected in this area.

We now describe the sensing infrastructure and social media data processing framework employed to collect experimental data.

Sensing Infrastructure

Each targeted terrain has been equipped with counting systems and Wi-Fi sensors, as depicted in Figure 5.3(b) and Figure 5.3(d). Counting systems ran computer vision algorithms on video feeds to count the numner of individual heads crossing a pre-defined cross-section in the street. The counting system provided every minute flow measurements in both directions (inflow and outflow), and had an accuracy of 92%-98%, depending on density conditions (Yuan et al., 2016). Wi-Fi sensors detected the presence of mobile devices located in their proximity (Yuan et al., 2016). For each device, the sensor hashed and stored its identifier, as well as its first and last detection time. We estimated that about one third of the counts from counting systems were identified by Wi-Fi sensors (Yuan et al., 2016). The matching rate between two adjacent Wi-Fi sensors was 3% - 4% of the total flow at the cross-section (Yuan et al., 2016).

The *Veemkade* terrain in the Sail and the *Zuidplein* terrain during Kingsday featured a counting system and a Wi-Fi sensor for both considered boundaries. Other terrains had only a single boundary equipped with both sensing devices. Table 5.1 lists counting systems and Wi-Fi sensors for each terrain. For boundaries without counting systems, the number of attendees traversing the cross-section (and the related flow information) has been estimated from Wi-Fi sensors, using the counting-to-Wi-Fi ratio calculated from the other boundary. A previous study on pedestrian traffic monitoring (Yuan et al., 2016) shows that the Wi-Fi counts which have been validated by the data from the counting sensors are known to have a 92% accuracy in high density conditions and 98% accuracy in low density conditions. Therefore, we consider the data is reliable for the purpose of our study.

Social Media Data Collecting & Processing Framework

We employed SocialGlass (Bocconi et al., 2015; Psyllidis et al., 2015b), an existing social media retrieval and enrichment framework, to listen from Twitter and Instagram streams for geo-located posts created within the city of Amsterdam during the first four days of SAIL 2015; for Kingsday 2016, we included the day of the event but also the previous and following days, for a total of 3 days of observation. We included in the analysis only geo-located posts, to maximise the spatial accuracy of the retrieved social media data. The inclusion of posts that are not geo-localised but related to the event (and, therefore, potentially localisable) is left to future work.

For each post, the latitude, longitude, timestamp, content, as well as the user id, are collected and stored in a database for further filtering and aggregation. Then, a *density estimation* module assigned each post to a targeted event terrain. Given as input a shape-file of the terrains, the module assesses the time and location of each post and user for each density estimation strategy. With $\widehat{k_1}$ and $\widehat{k_2}$, posts were assigned according to the geo-boundaries of the terrains. With $\widehat{k_3}$ and $\widehat{k_4}$, posts were assigned according to the geo-boundaries of possible routes that could lead to the terrains.

Table 5.2 reports descriptive statistics about the number of geo-located posts and unique users identified for terrains during the two events. A manual inspection of all the posts from the event terrain showed that a high percentage of them referenced the event (Yang et al., 2016).

The basic density estimation strategy $\widehat{k_1}$ captured a limited number of social media activities. This is to be expected, considering the generally low fraction of posts that are also geo-located – especially in Twitter, where geo-located posts are rare (around 1%) (Hecht et al., 2011). *Sumatrakade*, the less attractive terrain, featured the least number of posts. *Javakade* and

Table 5.2: Descriptive statistics of social media data captured by geo-, speed- and flow-based density estimation methods.

			Twitter		Instagram	
			#User	#Post	#User	#Post
$\widehat{k_1}$	Sail	Ruijterkade	16	24	25	28
		Sumatrakade	4	4	1	1
		Javakade	23	36	285	343
		Veemkade	6	22	61	86
	Kingsday	Zuidplein	2	2	4	4
$\widehat{k_2}$	Sail	Ruijterkade	19	31	38	45
		Sumatrakade	24	32	284	341
		Javakade	24	38	286	345
		Veemkade	10	28	76	104
	Kingsday	Zuidplein	4	5	20	21
$\widehat{k_3}/\widehat{k_4}$	Sail	Ruijterkade	349	717	2662	3577
		Sumatrakade	205	283	1113	1381
		Javakade	193	308	1026	1362
		Veemkade	466	877	3340	4554
	Kingsday	Zuidplein	191	355	3996	4925

Veemkade were the most popular, especially in terms of Instagram posts and users. This is also to be expected, given their proximity and access to tall ships and other points of interest. In Instagram, where geo-located posts are less sparse than in Twitter, *Ruijterkade* featured less posts than *Javakade* and *Veemkade*, indicating that attendees had less reasons to take pictures from that transit terrain. *Ruijterkade* has been comparably popular to *Javakade* and *Veemkade*; this is likely due to the proximity to the central station, a point of interest that attracts a lot of "check-in" posts from tourists and commuters. With other estimation strategies, the number of captured social media activities and users increases up to one order of magnitude, from more than 300 users to around 4000 users. *Sumatrakade* featured the largest relative increase, due to its close proximity to *Javakade*.

Comparison Metrics

Density values are compared with three metrics commonly used in time series analysis: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) (Hyndman & Koehler, 2006), and Pearson temporal correlation (Hauke & Kossowski, 2011). MAE measures the mean of absolute difference between two time series: a small distance would indicate similar time series in terms of magnitude. MAPE measures the mean of relative difference between two time series. The attendees density in a given event terrain greatly varies over time. Also, a city-scale event is not equally interesting through its whole duration. We therefore expect variations in the amount of attendees that feel compelled to share their experience on social media. Pearson temporal correlation computes the temporal correlation of two time series: a larger correlation would indicate the two time series have similar evolution patterns over time. The Pearson temporal correlation requires the time series data to follow a normal distribution (Hauke & Kossowski, 2011). We verified this condition for all density distributions using the Kolmogorov-Smirnov test (Plerou et al., 1999).

5.5 Results

This section presents and compares the density ($Persons/M^2$) estimation performance of the four considered methods. We first present the estimated densities; then, we assess their accuracy by comparing the calculated figures against density measured by the sensing infrastructure. Finally, we perform

Table 5.3: Density of people (#Persons/M^2) estimated by geo-, speed-, and flow-based estimation methods based on social media data, compared with sensor data.

		Sensor	$\widehat{k_1}$	$\widehat{k_2}$	$\widehat{k_3}$	$\widehat{k_4}$
Sail	Ruijterkade	2.269e-1	2.266e-5	2.639e-5	1.467e-3	2.935e-2
	Veemkade	2.586e-1	4.924e-5	5.762e-5	3.256e-3	2.606e-2
	Javakade	1.727e-1	1.384e-4	1.370e-4	7.495e-4	1.150e-1
	Sumatrakade	1.018e-1	7.246e-5	5.611e-4	2.803e-3	3.619e-2
Kingsday	Zuidplein	3.036e-1	1.302e-5	4.557e-5	8.998e-3	1.352e-1

Density of people estimated using social media is on hourly basis according to definition. 5.4

Density of people measured using sensor data is the mean of density in the same time window as the social media based methods calculated.

a sensitivity analysis on the Δt parameter of the speed- and flow- based models.

5.5.1 Results of Density Estimation

Density estimates and sensor measurement are calculated on an hourly basis. The technique used to process sensor measurements is described in previous work (Yuan et al., 2016). In $\widehat{k_3}$ and $\widehat{k_4}$ Δt is set to 30 minutes. Flow values in $\widehat{k_4}$ are obtained averaging, for each boundary, flow data produced during the Δt preceding the considered time window. Table 5.3 reports the density estimated by the four methods, and measured with sensors for the four SAIL terrains and the Kingsday terrain. Figure 5.4 shows for each of the considered terrains the temporal evolution of the estimated densities, to compare them with the density measured with sensors. Estimations from geo-based methods $\widehat{k_1}$ and $\widehat{k_2}$ are 3-4 orders of magnitude lower than density measured by sensors. This is due to the sparsity of social media data within the terrain areas, and in the considered time frame. That $\widehat{k_1}$ is performing worse in Zuidplein (for Kingsday) than on other terrains (for Sail) may be because the area of Zuidplein (0.64ha) is smaller than the area of other terrains (de Ruijterkade 6.12ha, Sumatrakade 1.38ha, Javakade 4.80ha, Veemkade 3.41ha). As social media data is sparse, the posts collected in the Zuidplein (on which $\widehat{k_1}$ is calculated based) is less than in other terrains.

Loosening the temporal and spatial constraints, $\widehat{k_3}$ estimates densities 2-

Table 5.4: Comparison between density measurement with sensor data and density estimates using geo- ($\widehat{k_1}$, $\widehat{k_2}$), speed- $\widehat{k_3}$, and flow-based $\widehat{k_4}$ methods.

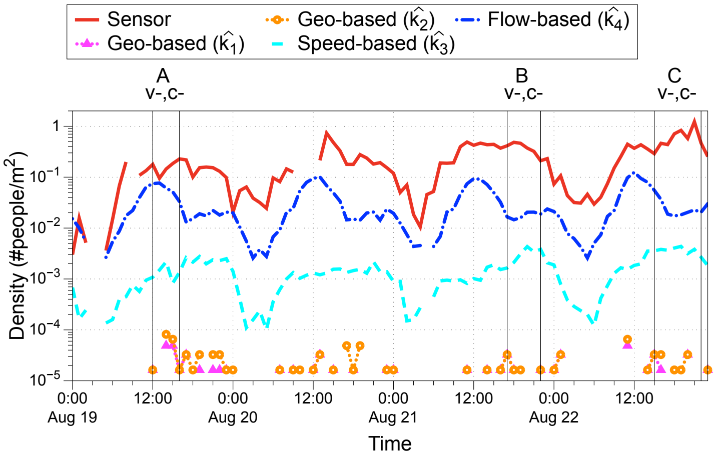
			$\widehat{k_1}$	$\widehat{k_2}$	$\widehat{k_3}$	$\widehat{k_4}$
MAE	Sail	Ruijterkade	.260	.281	.228	.119
		Veemkade	.353	.354	.258	.233
		Javakade	.218	.216	.173	.081
		Sumatrakade	.156	.114	.103	.068
	Kingsday	Zuidplein	.304	.304	.297	.162
MAPE	Sail	Ruijterkade	.9999	.9998	.9879	.8474
		Veemkade	.9995	.9992	.9588	.8569
		Javakade	.9962	.9962	.9705	.5667
		Sumatrakade	.9995	.9764	.9344	.7198
	Kingsday	Zuidplein	.9999	.9997	.9529	.5235
Spearman correlation	Sail	Ruijterkade	.083	-.094	.655***	.698***
		Veemkade	.031	.123	.634***	.537***
		Javakade	.296*	.308*	.486***	.690***
		Sumatrakade	NA	.378*	.586***	.695***
	Kingsday	Zuidplein	.086	.183	.596***	.864***

MAE: Mean Absolute Error.

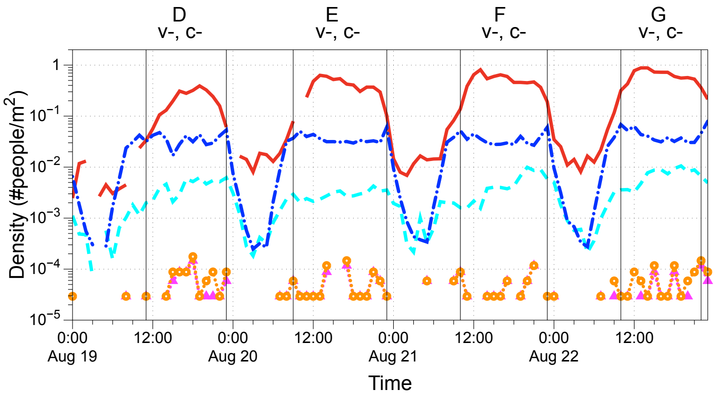
MAPE: Mean Absolute Percentage Error.

Spearman correlation marked with * and *** indicates p-value$<$ .05 and p-value$<$ .001, respectively.

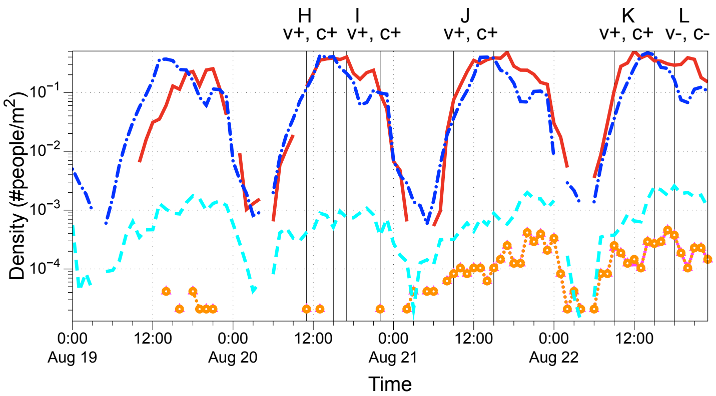
3 orders of magnitude lower than the densities measured with the sensing infrastructure. Finally, $\widehat{k_4}$, which uses flow information to scale the density estimated by $\widehat{k_3}$, reaches 1-2 magnitude orders lower than density measured with sensors. In the following, we discuss the results using metrics in more detail.



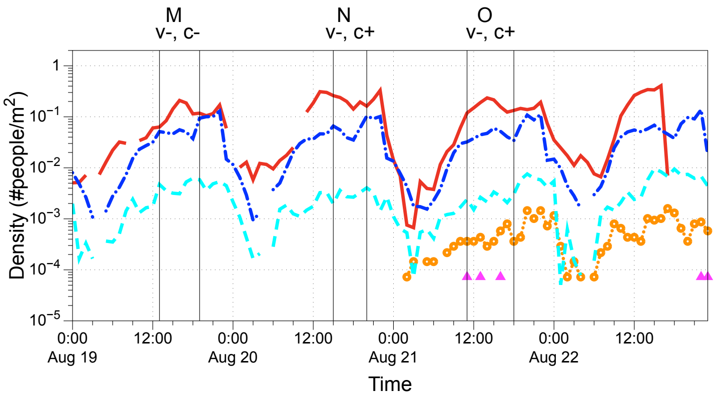
(a) Terrain 1: Ruijterkade



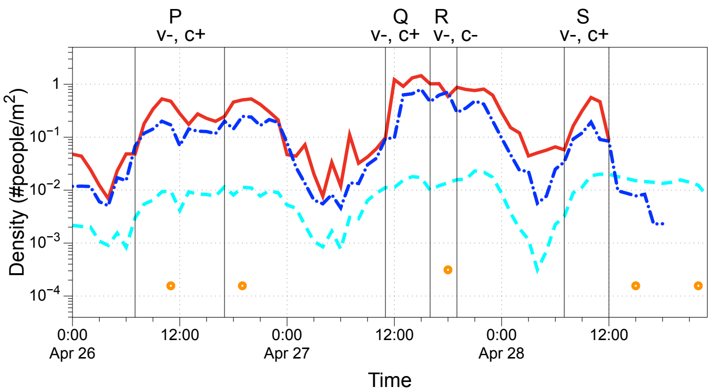
(b) Terrain 2: Veemkade



(c) Terrain 3: Javakade



(d) Terrain 4: Sumatrakade



(e) Kingsday, Terrain 1: Zuidplein. Sensor data is only available till 12:00 April 28

Figure 5.4: Evolution of density (P/m^2) estimates and sensor measurement during SAIL 2015 and Kingsday 2016. A to S denote the ID of selected periods which have similar or distinct value and temporal correlation listed in Table 5. "v+" denotes similar value. "v-" denotes distinct value. "c+" denotes similar temporal correlation. "c-" denotes distinct temporal correlation.

Mean Absolute (Percentage) Error

Table 5.4 (lines 2–11) reports the MAE and MAPE of each density estimation strategy, compared with measures based on sensor data. Geo-based density estimation methods $\widehat{k_1}$ and $\widehat{k_2}$ feature poor performance, with estimation errors up to 99%. The speed-based method $\widehat{k_3}$ provides slightly better performance, with an average 94% error. $\widehat{k_4}$ is the best in the pool, with an average error of 74%, decreasing to 56% in the *Javakade* terrain.

The relatively large absolute difference of $\widehat{k_3}$ and $\widehat{k_4}$ with regard to sensor data may be caused by the sparsity of the geo-located social media data, i.e. geo-located posts represent only a fraction of all the posts, especially in the Twitter platform (Hecht et al., 2011). What is more, social media have a relatively small penetration rate in the overall population.[8]

Spearman Temporal Correlation

The density measured with sensor data in Figure 5.4 shows daily patterns for all terrains in two cases, reaching a peak between 14:00 and 16:00, and minimum between midnight and 6:00.

Missing values are due to maintenance or disruptions with the sensing infrastructure. Density estimated with social media data, shows a distinct temporal pattern for each density estimation method.

In the following we analyse the performance of each method, by visually comparing the density curves in Figure 5.4, and by commenting on the Pearson temporal correlations shown in Table 5.4 (line 12-16). Due to sparsity issues, $\widehat{k_1}$ and $\widehat{k_2}$ fail to provide usable density estimates for all terrains, and in almost all time windows. The only exception is *Javakade*, where on August 21 and August 22 an increasing number of attendees active in social media allowed for a continuous density curve, but featuring a weak temporal correlation ($\widehat{k_1}$= .296, $\widehat{k_2}$= .308; p-value $< .05$) with sensor data.

The speed-based density estimation method ($\widehat{k_3}$) produces density estimates for most of the hourly time windows and for all terrains. $\widehat{k_3}$ features strong and significant temporal correlation with the sensor density time series. The result shows the benefits deriving from the consideration of attendees that could potentially be present in the terrains, but that post at locations within walking distance from the target event terrain.

The flow-based density estimation method $\widehat{k_4}$ achieves best results. Peak hours with $\widehat{k_4}$ fall into the same range of sensor measures. This could be explained by the scaling effect of flow data, an hypothesis supported by the relevant improvement in terms of temporal correlation (> 0.1) that can be

[8]Twitter, for instance, has a 17% reach in the Netherlands (source https://www.statista.com/statistics/279539/twitter-reach-in-selected-countries/).

observed from *Javakade* and *Sumatrakade*. However, there are also exceptions such as the correlation for the *Veemkade* terrain decreases (< 0.1).

Daily patterns could be observed in Figure 5.4 for each terrain, reaching the minimum between 14:00 and 18:00, and the maximum between 7:00 and 11:00. These peak hours differ from those of sensor data.

However, during the active hours (11:00-20:00) of event days the performance is varying in different cells and days. This is particularly obvious in *Veemkade*, where $\widehat{k_4}$ estimation shows plateau while sensor estimation reaches a peak in the afternoon.

Figure 5.4(e) shows the density estimation for the second case, Kingsday 2016, at terrain Zuidplein based on social media and sensor data. Similar to the first case, $\widehat{k_1}$ and $\widehat{k_2}$ fail to provide usable density estimation in all time windows. The speed-based density estimation method $\widehat{k_3}$ and flow-based density estimation method $\widehat{k_4}$ provide results for 3 days featuring strong and significant temporal correlation with the sensor density time series. They all clearly shows daily patterns during three days. $\widehat{k_4}$ featured better performance on both mean absolute percentage error and correlation compared with $\widehat{k_3}$ across all days. Density estimation by $\widehat{k_4}$ and sensor data on the second day (the day of the event) reaches the highest value among all three days, followed by the first day which is particularly active during the night. On the third day, $\widehat{k_3}$ features more stable estimation till the end of the day because the sensor data is only available till 12:00 on the third day, as such the $\widehat{k_4}$ is also affected by the lacking of flow information.

5.5.2 Δt Sensitivity Analysis

We now investigate how the performance of $\widehat{k_3}$ and $\widehat{k_4}$ density estimation methods changes with varying values of Δt, i.e. the model parameter controlling the temporal scope for micro-posts not created within a terrain of interest. We test values of Δt ranging from 5 minutes to 60 minutes, the length of the time window in this method. Results are shown in Figure 5.5. The $\widehat{k_4}$ method is robust to variations of Δt, although optimal performance is achieved for $\Delta t > 20$ minutes. With $\widehat{k_3}$, the temporal correlation of the density estimated in all terrains increases with increasing values of Δt, to stabilise between 30 minutes and 40 minutes. Interestingly, variations are not consistent across terrains. *Veemkade*, for instance, is most affected by changes in the Δt parameter, especially in terms of temporal correlation. On the other hand, estimates in *Ruijterkade* are the most robust. We believe that such inconsistent behaviour is due to differences in the properties of the terrains: *Ruijterkade* is a transit terrain, where attendees are less likely to stop during normal traffic conditions. Therefore, taking longer time frame into consideration does not significantly affect the amount of social media users accounted in the density calculation.

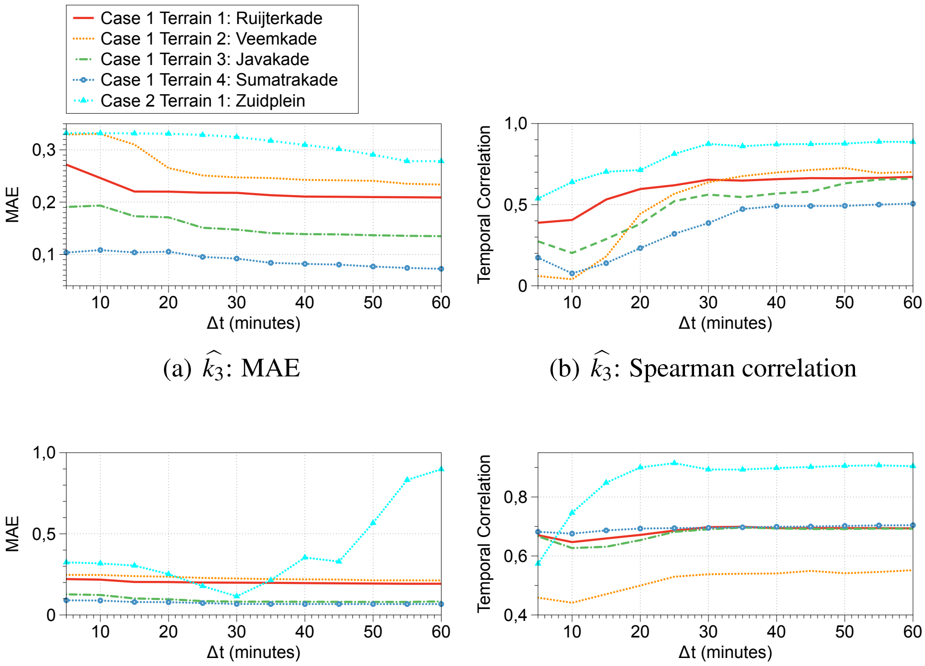
In the second case, the $\widehat{k_3}$ in Zuidplein is robust. However, the $\widehat{k_4}$ is not as robust as in terrains in the first case. It reaches the lowest mean absolute error when the value of Δt is around 30 minutes, then the mean absolute error is increased along with increasing of Δt, indicating that Zuidplein is more sensitive with regard to the variation of temporal scope. We account the result to the spatial characteristics of Zuidplein. As a pedestrian square, Zuidplein connects Amsterdam OUD-Zuid, CBD area and Station of Amsterdam Zuid, which is visited by a large number of people everyday. However, there are several other streets and roads which also connect these places and are in parallel with the Zuidplein, such as Eduard van Beinumstraat, Beethovenstraat and Parnassusweg. Therefore, loosing temporal and spatial constraints will easily introduce errors in calculating number of people who passed Zuidplein instead of other ways, which consequently increases errors in the density estimation.

5.6 Discussion

This section discusses the result of density estimation of each terrain in two cases. In order to get more insights about similar or distinct density estimations, we also look into several factors (e.g. temporal, demographic factors) and discuss their influences.

The $\widehat{k_2}$ in *Javakade* and *Sumatrakade* provide similar density estimation on Aug 21 and Aug 22, the weekend days. The improved performance in *Sumatrakade* with $\widehat{k_2}$ may be explained by the contiguity of the terrain with *Javakade*. It indicates that on social media the density estimation is sensitive to surroundings.

The daily patterns observed using $\widehat{k_4}$ from social media data and sensor data are different, which could be explained by the different types of activities captured by the two infrastructures – respectively, pedestrian movement and social media communication. Intuitively, some time slots during the event are more worthy of communication than others (e.g. ships during good lighting conditions, fireworks); on the other hand, the amount of atten-



(a) $\widehat{k_3}$: MAE (b) $\widehat{k_3}$: Spearman correlation

(c) $\widehat{k_4}$: MAE (d) $\widehat{k_4}$: Spearman correlation

Figure 5.5: The effects of Δt on the performance of speed-based density estimation method $\widehat{k_3}$ and $\widehat{k_4}$.

dees visiting schedules are affected by other factors (e.g. time and day of the week). However, some of communication oriented activities, such as Fireworks (lasting a maximum of 30 minutes) at 11pm each day in Sail event, are not captured by $\widehat{k_4}$, i.e. no peaks around 11pm on $\widehat{k_4}$. This may account for the influence of the length of time window selected for this experiment, i.e. 1h duration of time window may neutralize the high crowds during the fireworks and low flows towards the end of the hour. Thus, shorter time windows might capture these peaks.

Density estimation using social media featured higher performance in the second case than in the first case. This could be attributed to the diverse fingerprints of events and terrains as activities during Sail enhanced distinction of pedestrian movement and social media communication more than activities during Kingsday in those terrains.

Results also show that during active hours (07:00-23:00), density estimation performance varies for different terrains and events. In order to get more insights into them, we selected a set of periods which have either very similar or very distinct density estimation through social media data compared to sensor data ($\widehat{k_4}$, flow-based strategy) according to Mean Absolute Error and Spearman Temporal Correlation shown in Table 5.5. For each period we derived information from the crowd for various aspects such as demographic (i.e. Age, Gender), role of people with regard to the city (i.e. resident, local tourist, foreign tourist) and PoI preference of people, extracted through the SocialGlass system.

During Sail event the density estimation during periods of H, I, J and K in *Javakade* reaches best performance, i.e. similar value and similar temporal correlation. We found that the gender distribution derived from social media is more equal in these periods compared with other period in the same terrain (i.e. L), or periods in other terrains (e.g. A, D, N). Results points toward a relationship between the gender distribution of social media users and the performance of density estimation. However, this does not hold in the second case, where periods of P, Q and S reach a similar correlation while having less distinct values but the gender distribution does not show obvious patterns. This result suggests that other factors, such as type of events and location of the terrain, also play a role in the performance of our methods.

With regard to periods D, E, F and G in *Veemkade* which show huge distinctions in density estimation with regard to the sensor based method, we found that there are more male residents. Recent research (Yang et al., 2016) found that male and resident social media users are less active during city-scale events. Thus the reverse observation may indicate that the representativeness of social media data w.r.t. the reality is decreased. Consequently, the performance of density estimation based on social media data is affected. *Veemkade* is the narrowest terrain on the route of Orange Route connecting Amsterdam Central Station with *Javakade* and *Sumatrakade*, and it hosted restoration services and other Point of Interest, where people would stop, stand still, and block or hamper the flow of attendees. These may lead to the result that more people are detected by sensors rather than from social media. Consequently, the density of people detected from sensors and social media is in different value and correlation during these periods.

The selected periods A, B and C which show both distinct value and temporal correlation are from *Ruijterkade*. We found that during these periods there are more female foreigners active in social media, visiting PoIs such as Art & Entertainment, Food and Shop & Services in this terrain. However, the pattern of their influences is not clear.

Table 5.5: Selected periods with similar or distinct MAE. and Temproal Correlation in density estimation based on sensor and social media data

	Terrain	ID	Value	Temporal Correlation	Day	Period (hh-hh)
Sail (Aug. 2015)	Ruijterkade	A	-	-	19th	12-16
		B	-	-	21st	17-22
		C	-	-	22nd	15-23
	Veemkade	D	-	-	19th	11-23
		E	-	-	20th	09-23
		F	-	-	21st	10-23
		G	-	-	22nd	10-23
	Javakade	H	+	+	20th	11-15
		I	+	+	20th	17-22
		J	+	+	21st	09-15
		K	+	+	22nd	09-15
		L	-	-	22nd	18-23
	Sumatrakade	M	-	-	19th	13-19
		N	-	+	20th	15-20
		O	-	+	21st	11-18
Kingsday (Apr. 2016)	Zuidplein	P	-	+	19th	07-17
		Q	-	+	20th	11-16
		R	-	-	20th	16-19
		S	-	+	21st	07-12

ID: refers to the ID of periods shown in Figure 4.

Value: the value of estimated density.

Temporal Correlation: the temporal correlation of estimated density.

’+’: denotes the similar value or temporal correlation.

’-’: denotes the distinct value or temporal correlation.

Density estimations during periods N and O in *Sumatrakade* show similar temporal correlation but distinct value. We found that proportion of gender and role of people derived from social media in these periods show diverse values, but their patterns are not obvious, which is similar to the periods in *Zuidplein* in the second case.

In *Zuidplein*, density estimations in periods of P, Q and S show similar temporal correlation but distinct values, while period R shows both distinct temporal correlation and value. We found that the proportion of gender, role and the PoI preference of people are diverse during these periods. However, the pattern of their impacts is not obvious.

Above insights of the selected periods indicate that demographics, role, PoI preference of crowd, type of events, location of terrains as well as other factors may affect density estimation performance using social media. To fully understand their impacts, it calls for future work on factor analysis on density estimation performance based on social media data.

5.7 Conclusions

The density of attendees in an event terrain is an important measure of success and safety for city-scale events. In this paper we investigated the suitability of geo-referenced social media data produced during a city-scale event as a source for attendee density estimation. Social media have been used in a variety of contexts to analyse the amount of attendees at high temporal granularity, but low spatial granularity (e.g. city scale). However, due to the inherent geographical sparsity of geo-located social media data, the analysis of attendance at higher spatial granularity (e.g. street-scale) received less attention.

This paper proposes three density estimation strategies based on pedestrian traffic flow theory – respectively geo-, speed- and flow-based density estimation – that were successfully validated during city-scale events. When applied to geo-located social media sources for all strategies and additional flow data source for flow-based strategy, these strategies mitigate the spatial sparsity problem by considering traffic conditions (speed distribution and flow) to account for attendees that perform event-related social media activity outside an event terrain of interest. Thanks to a sophisticated sensing infrastructure deployed during SAIL 2015 and Kingsday 2016 in Amsterdam in the Netherlands, we assessed the performance of our methods on 5 event terrains characterised by different morphology and relevance to activities in both events. The flow-based method achieves promising performance in all terrains, both in terms of relative mean difference (from 20% to 120% improvement with regard to other methods) and temporal correlation (between .54 and .87). The speed-based method also features strong temporal correlation (between .49 and .65), but with higher estimation errors. Geo-based methods can yield useful results only when the amount of social media activity in the targeted terrain is sufficiently high.

We show that several factors play a significant role in terms of estimation accuracy and temporal correlation, such as the properties of a terrain, demographics, role and PoI preferences of the crowd. In Sail 2015, an attractive and trafficked terrain like *Veemkade* featured lower estimation accuracy and lower correlation than other terrains; a trafficked but less interesting terrain like *Ruijterkade* featured maximal temporal correlation but low estimation accuracy; a less trafficked terrain like *Javakade* featured higher estimation accuracy, but lower temporal correlation. Across all terrains, it is observed that maximal performance (i.e. higher temporal correlation and estimation precision than other terrains) is achieved with equal proportion of male and female in the crowd.

In the second case, Kingsday 2016, the trafficked terrain *Zuidplein* featured high correlation. The sensitivity analysis showed by loosing temporal and spatial constrains that the speed-based and flow-based methods achieve optimal performance when including users active at walking distance, and within 30-40 minutes from the temporal windows of observation. The characteristics of people counted for density estimation also affect the result. *Javakade* in the first case featured best performance with equally distributed gender of social media users than any other cells. Other factors, such as role and PoI preference of people, different types of events, also introduce influences on the result, but the patterns of their impacts are not clear.

The experimental result and the identification of influencing factors on the one hand help to avoid bias in applying this method for density estimation using social media, while on the other hand they call for future research in order to improve the estimation performance. In the next step we plan to take into consideration activity times of attendees, and investigate if the actual attendee speed distribution on the event terrain can be used for optimizing the density estimation. Further, we are going to zoom-in on the relation existing between traffic conditions and social media activity, to seek for stronger evidences of laws that relate attendees density with mobile online activity. We are also going to improve our estimation methods by using counting systems to provide speeds, using non-geo posts or posts with PoI information (e.g. from Facebook pages) in order to overcome data sparsity, using auto-filtering techniques to enhance posts filtering performance, and so on. We plan to compare the performance of our methods in various contexts of city-scale events, having different nature, size, and position in the city. Finally, we will explore the impact of factors, such as demographics, role, PoI preference of crowds, on the density estimation performance.

Chapter 6

Conclusions, implications and recommendations

In this chapter, we present our main findings and conclusions for each research question, followed by the overall conclusions and implications for practice. Finally, we provide recommendations for future research.

6.1 Main findings

The main objective of this dissertation research is to understand how social media can be used as a data source that provides meaningful information about the crowd for crowd management purposes. To achieve this objective, four research questions are answered in this dissertation thesis. In the following, the main findings of this research are structured in four parts corresponding to these four research questions in Section 1.3. In this section, we answer these research questions consecutively.

RQ1. To what extent are social media data able to characterize crowds in city events, in terms of demographic composition, city-role composition, Spatio-temporal distribution, Points of Interest preferences and word use? (Chapter 2)

To answer this research question, we selected a set of state-of-the-art methods for deriving information about crowds in terms of age, gender, city-role, crowd temporal distribution, post position, Points of Interest and word use, and perform a case study in two city events, i.e. Sail 2015 and King's Day 2016. The findings are as follows.

- **Age**: the age distribution during King's Day (SD: 0.201) is more evenly distributed than during the Sail event (SD: 0.213). This is in line with our expectation regarding the event participants' composition (i.e. the proportion of people in age groups of young, young-adult, adult and old) of the events' crowds according to the event programs: King's Day attracts families, while the Sail event attracts more (young) adults. This provides evidence that this information is reliable.

- **Gender**: the observed gender composition (Male/Female, Twitter: 1.3, Instagram: 0.7) from social media is similar on Sail and King's Day, which is not in line with the expected gender composition in these two events, i.e. more male visitors on Sail than King's Day. Only in sub-areas, such as the Javakade in Sail and the Zuidplein in King's Day, the gender of people observed in the former one is more equally distributed than the latter one, which may be caused by more recreation (bars and restaurants) and celebrities on Zuidplein, during King's Day give rise to the Instagram usage, which attracts more female users.

- **City-role**: less local tourists than residents and foreign travelers are observed during the King's Day (14.4%) event than during Sail (18.0%) from social media. This may be explained by the fact that people in other cities are more willing to travel to Amsterdam for a five-year festival Sail event, rather than the annual festival King's Day event.

- **Crowd temporal distribution**: the temporal distribution of crowd observed from social media during sub-events which take place in large areas, lasts for several hours during the daytime, e.g. the Sail-in parade and King's Day boat parade, is in line with the expected temporal distribution of crowd according to event programs.

- **Post Position and Points of Interest**: more social media usage and visited Points of Interest are observed in the IJhaven area during Sail than during the King's Day event. This is in line with the event programs that activities take place during Sail are around the IJhaven area where ships docked, while activities take place during King's Day are distributed in the whole city.

- **Word use**: words about event topics and people's emotions in different events are successfully captured in the top words used in social media. In the meantime, people are more willing to share their negative emotions, such as crowded perceptions, on Twitter.

So based on these findings above, we conclude that social media data can be used for deriving different types of information about crowds to different extents. With respect to demographic and city-role, the age composition (the proportion of people in age groups of young, young-adult, adult and old) and city-role composition (the proportion of people in groups of resident, local traveler, and foreign traveler) of crowds derived from social media data is in line with the expectation according to event programs, and hence seems plausible. However, gender composition, namely the proportion of male and female, derived from social media data is not in line with expectations according to event programs and organizers. Moreover, the variance of temporal and spatial distribution (post position and PoI) of crowds is in line with the expectation according to event programs. Furthermore, the top words captured on social media during events are in line with the event characteristics, such as event topics and expected emotion from crowds.

RQ2. To what extent are social media data able to estimate the sentiment of crowds in city events? (Chapter 3)

To answer this research question, we performed a sentiment analysis using social media data in several (city) events. We did so by constructing a novel sentiment annotated dataset consisting of social media texts collected during diverse city events. A set of selected state-of-the-art methods are validated using the constructed dataset.

Our findings show that machine learning based (ML-based) methods show better performance than lexicon-based methods in most situations in terms of sentiment estimation. In particular, the ML-based method Linear Support Vector Classifier (LinearSVC) reaches the minimal estimation error at approximately 0.177 when trained and tested with an event-based dataset constructed using social media data collected during city events. Because of these, we recommend using the ML-based method LinearSVC, trained with event-based data, to predict sentiments in a crowd using social media.

Based on these findings, we conclude that social media data can be used for sentiment analysis using machine learning techniques, which reaches meaningful results as it correctly identified sentiment in 82.3% of social media posts. These results are encouraging, and suggest that social media data could be used to help crowd managers better understand crowds attending city events.

RQ3. To what extent are social media images able to count people in city events? (Chapter 4)

To answer this research question, we investigated the effectiveness of four state-of-the-art methods on estimating crowd size using social media images in city events. We also investigated the impact of image characteristics on the estimation performance for different methods. We selected four state-of-the-art methods after performing a comprehensive literature review and used the crowd size and image characteristics annotated dataset to examine the selected methods.

We found that social media images can be used to estimate the crowd size in the image. Findings show that the so-called direct method [1], i.e. Darknet Yolo (Redmon et al., 2016; Redmon & Farhadi, 2017), reaches better estimation accuracy than so-called indirect methods [2], i.e. Cascade methods A and B (Sindagi & Patel, 2017). Specifically, Darknet Yolo reaches the highest accuracy (72%) in crowd size level estimation and the highest accuracy (38%) in estimating the specific number of people when less than twenty people appear in an image. At the same time, the findings about image characteristics impact on methods performance indicate that social media selfie pictures in parallel view [3] with selfie people in full face [4] and gatherings in fixed distributed are more feasible to Faceplusplus and Darknet Yolo for crowd size estimation, while pictures taken from top view with gatherings in gradient distribution are more suitable to Cascaded methods. We recommend using the Darknet Yolo method to estimate and predict the crowd size

[1] Direct methods are methods that identify persons using handcrafted features (e.g. face, head, shoulder and legs of people) in an image and accumulate them as the number of people in an image.

[2] Indirect methods are methods that count people using non-handcrafted features (learned features) applied with learning algorithms or statistical analyses.

[3] The parallel view refers to the view that the image is captured from the camera. An image is annotated as 'parallel' if it is captured by a camera in parallel, rather than from top (annotated as 'top') or from the view between top and parallel (annotated as 'between').

[4] Full face is a value of the image characteristic face types. A selfie image may contain faces in full faces, partial faces, back faces, mix faces or none faces.

in city events based on social media images, in terms of levels of crowd size as well as the specific number of people if it is the low dense environment.

RQ4. To what extent are social media data able to estimate the density of people in city events? (Chapter 5)

To answer the final research question, we proposed a new method to estimate density of people using social media data in city events based on pedestrian traffic flow theory. It consists of three estimation strategies, i.e. geo-based, speed-based and flow-based density estimation strategies, that were validated during two city events, i.e. Sail 2015 and King's Day 2016. The main findings are as follows.

We show that the proposed density estimation methods using three different strategies, i.e. geo-, speed- and flow-based, yield different estimation accuracy. The flow-based density estimation method achieves promising performance in all monitoring areas, both in terms of relative mean difference (from 20% to 120% improvement with regard to other methods) and temporal correlation (between 0.54 and 0.87). The speed-based method also features a strong temporal correlation (between 0.49 and 0.65), but with higher estimation errors. Geo-based methods can yield useful results only when the amount of social media activity in the targeted monitoring area is sufficiently high. In the meantime, several factors play a significant role in terms of estimation accuracy and temporal correlation, such as the properties of a monitoring area, demographics, role, and Points of Interest preferences of the crowd. If across all monitoring areas, the maximal performance (i.e. higher temporal correlation and estimation precision than other areas) is achieved with equal proportion of male and female in the crowd. The sensitivity analysis showed that the speed-based and flow-based methods achieve optimal performance when including active users at walking distance, and within 30-40 minutes from the temporal windows of observation.

6.2 Overall conclusions

Based on the findings, the following conclusions can be drawn: first, this research provides evidence showing that information about crowds derived from social media, entailing age, city-role, spatio-temporal distribution, Points of Interest preference, word use, sentiment, crowd size and density estimation is in line with the estimates of the ground truth during city events stemming from events programs. In the meantime, the accuracy and reliability of the derived information from social media is influenced by event properties, attendee profiles, and location characteristics.

Further, using machine learning based method Linear Support Vector Classifier (LinearSVC), trained with the constructed event-based social media sentiment dataset, it is possible to estimate the sentiments of people in the crowd during city events. The Darknet Yolo method can effectively estimate the size of crowd from social media images during city events. Moreover, we can conclude that, based on pedestrian traffic flow theory, social media data can be used to formulate a new method to estimate the density of people in the crowd.

When it comes to crowd management applications, in the event planning phase, information about crowd derived from social media collected in history events can be used for three purposes, i.e. (a) crowd simulation, where the historical data is used for simulating the dynamics of crowds in terms of their movement, Spatio-temporal distribution, density, and so on, (b) inferring crowd management guidelines, such as strategies to deal with most likely crowded time and area, such as queues and bottlenecks, and (c) preparing what-if scenarios, i.e. identify scenarios such as event build-up and break-up phases of the event, and predefine crowd management measures if these scenarios occur. The sparsity of geo-referenced social media data which may affect deriving information about crowds can be eliminated by scaling up the number of geo-referenced social media data based on probabilistic strategies discussed in Chapter 5. The bias of social media usage in city events can be reduced (Culotta, 2014; Kuru & Pasek, 2016; Firth, 1993) based on the insights of case studies.

In the operational phase of crowd management (e.g. real-time monitoring and intervention deployment), the derived (near) current information about crowds using social media data can be used to monitor crowds, predict accidents, and apply feasible predefined measures. Crowd managers keep monitoring the crowds through derived information. Once the information, such as density or sentiment of crowds, reaches predefined critical levels, crowd managers can apply predefined measures to manage the crowds. As the process of deriving information from social media is not in real-time, the latency between derived information and the reality may affect the crowd management decision making, i.e. the shorter latency the higher utility of derived information for crowd management. Case studies in our research shows the latency of deriving information about crowds using social media data can be shortened to 3 to 5 minutes using enhanced computational power and network bandwidth in data processing.

To summarize, this research provides insights into data and methods about deriving information of crowds using social media in city events for crowd management. A new method to estimate the density of people in the crowd using social media has been proposed, validated and verified.

6.3 Implications for crowd management practice

The findings and conclusions in this dissertation provide important implications for practice.

6.3.1 Adopt Web-based data sources in the crowd management operations

Current crowd management best-practices suggest the adoption of a combination of data sources, from stewards on the event terrain regularly reporting the current status, to continuous feeds coming from ICT infrastructures (Li, 2019). As discussed in the introductory chapter, existing solutions for deriving data of crowds has advantages and disadvantages in terms of adoption, precision, semantic richness, and cost. Our work (Chapter 2,3,4,5) shows that Web-based communication channels, specifically social media, can be used to provide qualitative and quantitative information that can be useful for crowd managers. As there are limitations also with social media data sources (e.g. data sparsity and bias), our recommendation is to consider them as complementary data sources in support to crowd managers' decision making.

6.3.2 Insight into biases introduced by social media sources and/or content analysis techniques

In the context of the two case studies, this research shows that social media data exhibits biases in various aspects, such as age, gender, city-role, and word use. For instance, a social media platform like Instagram is more popular with teenagers than adults; despite the more balanced demographic distribution of attendees during events, social media data collected from city-scale events might reflect the bias present in the general platform population. Content analysis methods can also introduce biases that are intrinsic to the business logic of such technique, or are dependent on the properties of the analysed data. For instance, a machine learning gender recognition algorithm might be biased to perform better for some demographics (e.g. due to the employed training data); at the same time, such algorithms could be of little to no use if the processed data (e.g. social media profile pictures) are not truthful. While bias reduction techniques exist (Culotta, 2014; Kuru & Pasek, 2016; Firth, 1993), we acknowledge that, to some extent, some form of bias is unavoidable. We therefore advocate for a responsible usage of social media data for crowd management purposes, where their value and utility is interpreted through the lenses of a good understanding of their biases in the context of their application. At the sametime, it is also possible to use the social media data for crowd management as a complement for a cost-effective system where the number of unbiased sensors (e.g. counting-systems) can be reduced in combining them with social media data.

6.3.3 Promote social media usage in city events for crowd management

Obviously, the utility of social media data for crowd management purposes is conditional to their existence in the context of the monitored event. Our research shows that, despite the presence of sparsity issues, publicly available social media content could be sufficient and very useful for providing quantitative and qualitative information about city-scale events. To reduce such sparsity, event organisers should promote the use of social media by the event's attendees, e.g. by means of advertising campaigns, or public challenges and competitions. In doing so, we stress the importance of openness and transparency towards their crowd management purposes, to allow conscious and informed data creation.

6.4 Recommendations for future research

The work presented in this thesis represents a substantial step towards a better understanding of the value and utility of social media data for crowd management planning and operations purposes. Based on the reported observations and findings, we elaborate on several opportunities for future work.

6.4.1 Address geo-referenced social media data sparsity

Geo-referenced social media posts, i.e. posts explicitly containing the geographical coordinates associated with the posting location, are important for deriving information about the amount and distribution of attendees in an event area. In our research, we apply probabilistic strategies to scale up the number of geo-referenced posts based on estimated ground truth data. To improve the amount of available geo-located posts, suitable existing geo-localisation techniques for non-georeferenced social media posts should be investigated. Geolocalization techniques allow for estimating the location of non-georeferenced social media posts based on various types of information contained in the posts, such as topics, timestamp, and user profile. Existing techniques (Paule et al., 2019; Middleton et al., 2018) allows the estimation of the location of non-georeferenced posts at different spatial granularity, e.g. street level, block level, and city level. By building on existing approaches (Paule et al., 2019; Middleton et al., 2018), future work should investigate the effectiveness of such geolocalization techniques in the context of social media posts created during city events.

6.4.2 Novel data-fusion techniques

Conventional solutions derive information about crowds based on traditional data sources such as stewards and ICT infrastructure, e.g. camera, counting-system, GPS tracker, Wi-Fi sensors. In our research, we use new data sources, the social media data, to derive such information. In particular, we proposed a new method to estimate density of crowds in city events using geo-, speed- and flow-based strategies. Indeed, the different types of data sources must not be seen as mutually exclusive. Future research can investigate to what extent the effectiveness of deriving information about crowds can be increased by fusing social media data with traditional data sources. This thesis provides an example. In Chapter 5 we describe a flow-based strategy, that uses pedestrian flow data collected from counting-systems as a prior knowledge for a crowd density estimation model based on social media data; the fusion, in this case, led to a 20% to 120% improvement in terms of relative mean difference with regard to other strategies without using pedestrian flow data for calibration. This approach could indeed be pushed forward, considering other types of data sources. For instance, data from GPS trackers in local areas can help with calibrating the parameters of methods estimating density of people in broader areas. Considering the problem of crowd sentiment estimation addressed in Chapter 3, videos and images captured from cameras in localised monitoring areas may be used to inform the estimation of the demographics and sentiment of crowds in broader areas using social media data.

6.4.3 Gaining experience by applications during running events

The information about the crowd extracted from social media content can find application in two distinct phases of a crowd management effort. In the event planning phase, the information can be used for analytics and simulation purposes. In the operational phase, it can be used for assessing the current situation for decision making. Our work had been designed and conducted with both phases in mind. Future work should investigate the impact that accuracy, reliability, and latency of information could have in the context of real planning and operational phases.

Bibliography

Abbasi, A., T. H. Rashidi, M. Maghrebi, S. T. Waller (2015) Utilising location based social media in travel survey methods: bringing twitter data into the play, in: *Proceedings of the 8th ACM SIGSPATIAL international workshop on location-based social networks*, ACM, p. 1.

Abbott, J., M. W. Geddie (2000) Event and venue management: Minimizing liability through effective crowd management techniques, *Event Management*, 6(4), pp. 259–270.

Abdul-Mageed, M., M. Diab, S. Kübler (2014) Samar: Subjectivity and sentiment analysis for arabic social media, *Computer Speech & Language*, 28(1), pp. 20–37.

Aiken, M., S. Balan (2011) An analysis of google translate accuracy, *Translation journal*, 16(2), pp. 1–3.

Al-Azani, S., E.-S. M. El-Alfy (2017) Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short arabic text, *Procedia Computer Science*, 109, pp. 359–366.

Al-Zahrani, M. S., N. F. Bissada, E. A. Borawski (2003) Obesity and periodontal disease in young, middle-aged, and older adults, *Journal of periodontology*, 74(5), pp. 610–615.

Alkhatib, M., M. El Barachi, K. Shaalan (2019) An arabic social media based framework for incidents and events monitoring in smart cities, *Journal of Cleaner Production*, 220, pp. 771–785.

Amerini, I., C.-T. Li, R. Caldelli (2019) Social network identification through image classification with convolutional neural network, *IEEE access*, 7, pp. 35264–35273.

Aston, N., T. Munson, J. Liddle, G. Hartshaw, D. Livingston, W. Hu (2014) Sentiment analysis on the social networks using stream algorithms, *Journal of Data Analysis and Information Processing*, 2(02), p. 60.

Aydoğan, E., M. A. Akcayol (2016) A comprehensive survey for sentiment analysis tasks using machine learning techniques, in: *INnovations in Intelligent SysTems and Applications (INISTA), 2016 International Symposium on*, IEEE, pp. 1–7.

Baccianella, S., A. Esuli, F. Sebastiani (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining., in: *Lrec*, vol. 10, pp. 2200–2204.

Balduini, M., A. Bozzon, E. Della Valle, Y. Huang, G.-J. Houben (2014) Recommending venues using continuous predictive social media analytics, *IEEE Internet Computing*, 18(5), pp. 28–35.

Balduini, M., E. Della Valle, D. Dell'Aglio, M. Tsytsarau, T. Palpanas, C. Confalonieri (2013) Social listening of city scale events using the streaming linked data framework, in: *International Semantic Web Conference*, Springer, pp. 1–16.

Berrigan, D., R. P. Troiano (2002) The association between urban form and physical activity in us adults, *American journal of preventive medicine*, 23(2), pp. 74–79.

Blanke, U., G. Tröster, T. Franke, P. Lukowicz (2014) Capturing crowd dynamics at large scale events using participatory gps-localization, in: *Intelligent Sensors, Sensor Networks and Information Processing, IEEE 9th International Conference on*, IEEE, pp. 1–7.

Blitzer, J., M. Dredze, F. Pereira (2007) Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification, in: *Proceedings of the 45th annual meeting of the association of computational linguistics*, pp. 440–447.

Bocconi, S., A. Bozzon, A. Psyllidis, C. Titos Bolivar, G.-J. Houben (2015) Social glass: A platform for urban analytics and decision-making through heterogeneous social data, in: *Proceedings of the 24th International Conference on World Wide Web*, ACM, pp. 175–178.

Boiy, E., M.-F. Moens (2009) A machine learning approach to sentiment analysis in multilingual web texts, *Information retrieval*, 12(5), pp. 526–558.

Botta, F., H. S. Moat, T. Preis (2015) Quantifying crowd size with mobile phone and twitter data, *Royal Society open science*, 2(5), p. 150162.

Bravo-Marquez, F., M. Mendoza, B. Poblete (2014) Meta-level sentiment models for big social data analysis, *Knowledge-Based Systems*, 69, pp. 86–99.

Cambria, E., D. Das, S. Bandyopadhyay, A. Feraco (2017) *A practical guide to sentiment analysis*, Springer.

Chan, A. B., Z.-S. J. Liang, N. Vasconcelos (2008) Privacy preserving crowd monitoring: Counting people without people models or tracking, in: *2008 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 1–7.

Chandra, S., A. K. Bharti (2013) Speed distribution curves for pedestrians during walking and crossing, *Procedia-Social and Behavioral Sciences*, 104, pp. 660–667.

Chen, J., G. Hsieh, J. U. Mahmud, J. Nichols (2014) Understanding individuals' personal values from social media word use, in: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, ACM, pp. 405–414.

Chen, K., S. Gong, T. Xiang, C. Change Loy (2013) Cumulative attribute space for age and crowd density estimation, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2467–2474.

Chen, K., C. C. Loy, S. Gong, T. Xiang (2012) Feature mining for localised crowd counting., in: *BMVC*, vol. 1, p. 3.

Cheng, Z., J. Caverlee, K. Lee, D. Z. Sui (2011) Exploring millions of footprints in location sharing services., *ICWSM*, 2011, pp. 81–88.

Costa, H., L. H. Merschmann, F. Barth, F. Benevenuto (2014) Pollution, badmouthing, and local marketing: The underground of location-based social networks, *Information Sciences*, 279, pp. 123–137.

Cottrill, C., P. Gault, G. Yeboah, J. D. Nelson, J. Anable, T. Budd (2017) Tweeting transit: An examination of social media strategies for transport information management during a large event, *Transportation Research Part C: Emerging Technologies*, 77, pp. 421–432.

Cramér, H. (1999) *Mathematical methods of statistics*, vol. 9, Princeton university press.

Cranshaw, J., R. Schwartz, J. Hong, N. Sadeh (2012) The livehoods project: Utilizing social media to understand the dynamics of a city.

Cruz, F. L., J. A. Troyano, F. Enríquez, F. J. Ortega, C. G. Vallejo (2013) 'long autonomy or long delay?'the importance of domain in opinion mining, *Expert Systems with Applications*, 40(8), pp. 3174–3184.

Culotta, A. (2014) Reducing sampling bias in social media data for county health inference, in: *Joint Statistical Meetings Proceedings*, pp. 1–12.

Daamen, W. (2004) Modelling passenger flows in public transport facilities.

Daamen, W., S. Hoogendoorn, P. Bovy (2005) First-order pedestrian traffic flow theory, *Transportation Research Record: Journal of the Transportation Research Board*, (1934), pp. 43–52.

Daamen, W., Y. Yuan, D. Duives, S. Hoogendoorn (2016) Comparing three types of real-time data collection techniques: Counting cameras, wi-fi sensors and gps trackers, *Proceedings of Pedestrian and Evacuation Dynamics 2016*.

Daganzo, C., Carlos (1997) *Fundamentals of transportation and traffic operations*, vol. 30, Pergamon Oxford.

Davies, A. C., J. H. Yin, S. A. Velastin (1995) Crowd monitoring using image processing, *Electronics & Communication Engineering Journal*, 7(1), pp. 37–47.

Davis Jr, C. A., G. L. Pappa, D. R. R. de Oliveira, F. de L Arcanjo (2011) Inferring the location of twitter messages based on user relationships, *Transactions in GIS*, 15(6), pp. 735–751.

Díaz, M., I. Johnson, A. Lazar, A. M. Piper, D. Gergle (2018) Addressing age-related bias in sentiment analysis, in: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.

Dickinson, G. (2019) How can we put a stop to the damaging rise of the insta-tourist?, URL https://www.telegraph.co.uk/travel/comment/instagram-tourists-geotag/.

Dreiseitl, S., L. Ohno-Machado (2002) Logistic regression and artificial neural network classification models: a methodology review, *Journal of biomedical informatics*, 35(5-6), pp. 352–359.

Duggan, M., J. Brenner (2013) *The demographics of social media users, 2012*, vol. 14, Pew Research Center's Internet & American Life Project Washington, DC.

Duives, D. C., W. Daamen, S. P. Hoogendoorn (2015) Quantification of the level of crowdedness for pedestrian movements, *Physica A: Statistical Mechanics and its Applications*, 427, pp. 162–180.

Duives, D. C., W. Daamen, S. P. Hoogendoorn (2016) The influence of the interaction characteristics on the movement dynamics of pedestrians, in: *Proceedings of the 8th International Conference on Pedestrian and Evacuation Dynamics (PED2016)*, University of Science and Technology of China Press.

Earl, C., E. Parker, A. Tatrai, M. Capra, et al. (2004) Influences on crowd behaviour at outdoor music festivals, *Environmental Health*, 4(2), p. 55.

Eash, R. (1999) Destination and mode choice models for nonmotorized travel, *Transportation Research Record: Journal of the Transportation Research Board*, (1674), pp. 1–8.

Ester, M., H.-P. Kriegel, J. Sander, X. Xu, et al. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise., in: *Kdd*, vol. 96, pp. 226–231.

Fang, Z., J. Yuan, Y.-C. Wang, S. M. Lo (2008) Survey of pedestrian movement and development of a crowd dynamics model, *Fire safety journal*, 43(6), pp. 459–465.

Favaretto, R. M., L. L. Dihl, S. R. Musse (2016) Detecting crowd features in video sequences, in: *Graphics, Patterns and Images (SIBGRAPI), 2016 29th SIBGRAPI Conference on*, IEEE, pp. 201–208.

Feng, J., L. D. A. Barbosa, V. Torres (2019) Systems and methods for social media data mining, uS Patent 10,496,654.

Firth, D. (1993) Bias reduction of maximum likelihood estimates, *Biometrika*, 80(1), pp. 27–38.

Florez, J., J. Muniz, L. Portugal (2014) Pedestrian quality of service: Lessons from maracanã stadium, *Procedia-Social and Behavioral Sciences*, 160, pp. 130–139.

Fruin, J. J. (1971) *Designing for pedestrians: A level-of-service concept*, HS-011 999.

Gao, S. (2015) Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age, *Spatial Cognition & Computation*, 15(2), pp. 86–114.

Georgiev, P., A. Noulas, C. Mascolo (2014) The call of the crowd: Event participation in location-based social services, AAAI.

Ghani, N. A., S. Hamid, I. A. T. Hashem, E. Ahmed (2019) Social media big data analytics: A survey, *Computers in Human Behavior*, 101, pp. 417–428.

Gilbert, C. H. E. (2014) Vader: A parsimonious rule-based model for sentiment analysis of social media text, in: *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf.*

Gong, V. X., W. Daamen, A. Bozzon, S. Hoogendoorn (2018a) Crowd characterization using social media data in city-scale events for crowd management, Tech. rep.

Gong, V. X., W. Daamen, A. Bozzon, S. P. Hoogendoorn (2019) Estimate sentiment of crowds from social media during city events, *Transportation Research Record*, p. 0361198119846461.

Gong, V. X., J. Yang, W. Daamen, A. Bozzon, S. Hoogendoorn, G.-J. Houben (2018b) Using social media for attendees density estimation in city-scale events, *IEEE Access*, 6, pp. 36325–36340.

Gong, X. (2016) Exploring human activity patterns across cities through social media data.

Guo, Z. (2009) Does the pedestrian environment affect the utility of walking? a case of path choice in downtown boston, *Transportation Research Part D: Transport and Environment*, 14(5), pp. 343–352.

Gurkhe, D., N. Pal, R. Bhatia (2014) Effective sentiment analysis of social media datasets using naive bayesian classification, *International Journal of Computer Applications (0975 8887)*, 99(13).

Guyon, I. (1997) A scaling law for the validation-set training-set size ratio, *AT&T Bell Laboratories*, pp. 1–11.

Han, Q., B. G. Dellaert, W. F. Van Raaij, H. J. Timmermans (2010) Visitors' strategic anticipation of crowding in scarce recreational resources, *Journal of Retailing and Consumer Services*, 17(6), pp. 449–456.

Handy, S. (1996) Urban form and pedestrian choices: study of austin neighborhoods, *Transportation Research Record: Journal of the Transportation Research Board*, (1552), pp. 135–144.

Hasan, S., X. Zhan, S. V. Ukkusuri (2013) Understanding urban human activity and mobility patterns using large-scale location-based data from online social media, in: *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, ACM, p. 6.

Hauke, J., T. Kossowski (2011) Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data, *Quaestiones geographicae*, 30(2), pp. 87–93.

Hawelka, B., I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, C. Ratti (2014) Geo-located twitter as proxy for global mobility patterns, *Cartography and Geographic Information Science*, 41(3), pp. 260–271.

Hecht, B., L. Hong, B. Suh, E. H. Chi (2011) Tweets from justin bieber's heart: the dynamics of the location field in user profiles, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 237–246.

Helbing, D., L. Buzna, A. Johansson, T. Werner (2005) Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions, *Transportation Science*, 39(1), pp. 1–24.

Hochmair, H. H., L. Juhász, S. Cvetojevic (2018) Data quality of points of interest in selected mapping and social media platforms, in: *LBS 2018: 14th International Conference on Location Based Services*, Springer, pp. 293–313.

Holson, L. M. (2018) Is geotagging on instagram ruining natural wonders? some say yes, URL https://www.nytimes.com/2018/11/29/travel/instagram-geotagging-environment.html.

Hoogendoorn, S. P., P. H. Bovy (2005) Pedestrian travel behavior modeling, *Networks and Spatial Economics*, 5(2), pp. 193–216.

Hu, M., B. Liu (2004) Mining and summarizing customer reviews, in: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 168–177.

Hyndman, R. J., A. B. Koehler (2006) Another look at measures of forecast accuracy, *International journal of forecasting*, 22(4), pp. 679–688.

Idrees, H., I. Saleemi, C. Seibert, M. Shah (2013) Multi-source multi-scale counting in extremely dense crowd images, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2547–2554.

Jamil, S., A. Basalamah, A. Lbath, M. Youssef (2015) Hybrid participatory sensing for analyzing group dynamics in the largest annual religious gathering, in: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, pp. 547–558.

Jiang, L., M. Yu, M. Zhou, X. Liu, T. Zhao (2011) Target-dependent twitter sentiment classification, in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, Association for Computational Linguistics, pp. 151–160.

Jiang, M., J. Huang, X. Wang, J. Tang, C. Wu (2014) An approach for crowd density and crowd size estimation., *JSW*, 9(3), pp. 757–762.

Kim, S., S. Park, J. S. Lee (2014) Meso-or micro-scale? environmental factors influencing pedestrian satisfaction, *Transportation Research Part D: Transport and Environment*, 30, pp. 10–20.

Kiritchenko, S., X. Zhu, S. M. Mohammad (2014) Sentiment analysis of short informal texts, *Journal of Artificial Intelligence Research*, 50, pp. 723–762.

Kohavi, R., et al. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Ijcai*, vol. 14, Montreal, Canada, pp. 1137–1145.

Kouloumpis, E., T. Wilson, J. D. Moore (2011) Twitter sentiment analysis: The good the bad and the omg!, *Icwsm*, 11(538-541), p. 164.

Krueger, R., Q. Han, N. Ivanov, S. Mahtal, D. Thom, H. Pfister, T. Ertl (2019) Bird's-eye-large-scale visual analytics of city dynamics using social location data, in: *Computer Graphics Forum*, vol. 38, Wiley Online Library, pp. 595–607.

Kuru, O., J. Pasek (2016) Improving social media measurement in surveys: Avoiding acquiescence bias in facebook research, *Computers in Human Behavior*, 57, pp. 82–92.

Lansley, G., P. Longley (2016) Deriving age and gender from forenames for consumer analytics, *Journal of Retailing and Consumer Services*, 30, pp. 271–278.

Lee, R., K. Sumiya (2010) Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection, in: *Proceedings of the 2nd ACM SIGSPATIAL international workshop on location based social networks*, ACM, pp. 1–10.

Lempitsky, V., A. Zisserman (2010) Learning to count objects in images, in: *Advances in neural information processing systems*, pp. 1324–1332.

Li, J. (2019) *Crowds inside out: Understanding crowds from the perspective of individual crowd members' experiences*, Ph.D. thesis, Delft University of Technology.

Li, Y.-M., T.-Y. Li (2013) Deriving market intelligence from microblogs, *Decision Support Systems*, 55(1), pp. 206–217.

Liang, Y., J. Caverlee, Z. Cheng, K. Y. Kamath (2013) How big is the crowd?: event and location based population modeling in social media, in: *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, ACM, pp. 99–108.

Lima, A. C. E., L. N. de Castro, J. M. Corchado (2015) A polarity analysis framework for twitter messages, *Applied Mathematics and Computation*, 270, pp. 756–767.

Lingad, J., S. Karimi, J. Yin (2013) Location extraction from disaster-related microblogs, in: *Proceedings of the 22nd international conference on world wide web*, ACM, pp. 1017–1020.

Liu, B. (2012) Sentiment analysis and opinion mining, *Synthesis lectures on human language technologies*, 5(1), pp. 1–167.

Longley, P. A., M. Adnan, G. Lansley (2015) The geotemporal demographics of twitter usage, *Environment and Planning A*, 47(2), pp. 465–484.

Luo, D., L. Bonnetain, O. Cats, H. van Lint (2018) Constructing spatiotemporal load profiles of transit vehicles with multiple data sources, *Transportation Research Record*, p. 0361198118781166.

Maas, A. L., R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, C. Potts (2011) Learning word vectors for sentiment analysis, in: *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1*, Association for Computational Linguistics, pp. 142–150.

Maley, D. W., R. R. Weinberger (2011) Food shopping in the urban environment: Parking supply, destination choice, and mode choice, in: *90th Annual Meeting of the Transportation Research Board, Washington, DC*.

Marana, A. N., S. A. Velastin, L. d. F. Costa, R. Lotufo (1998) Automatic estimation of crowd density using texture, *Safety Science*, 28(3), pp. 165–175.

Marsden, M., K. McGuinness, S. Little, N. E. O'Connor (2017) Resnetcrowd: A residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification, in: *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, pp. 1–7.

Martella, C., J. Li, C. Conrado, A. Vermeeren (2017) On current crowd management practices and the need for increased situation awareness, prediction, and intervention, *Safety science*, 91, pp. 381–393.

Martin, A. (2006) *Factors influencing pedestrian safety: a literature review*, PPR241, TRL Wokingham, Berks.

Martínez Cámara, E., M. Valdivia, M. Teresa, J. M. Perea Ortega, L. A. Ureña López (2011) Técnicas de clasificación de opiniones aplicadas a un corpus en español.

May, A. D. (1990) *Traffic flow fundamentals*, Prentice-Hall.

McAuley, J., J. Leskovec (2012) Image labeling on a network: using social-network metadata for image classification, in: *European conference on computer vision*, Springer, pp. 828–841.

McCormack, G. R., A. Shiell (2011) In search of causality: a systematic review of the relationship between the built environment and physical activity among adults, *International Journal of Behavioral Nutrition and Physical Activity*, 8(1), p. 125.

Middleton, S. E., G. Kordopatis-Zilos, S. Papadopoulos, Y. Kompatsiaris (2018) Location extraction from social media: Geoparsing, location disambiguation, and geotagging, *ACM Transactions on Information Systems (TOIS)*, 36(4), p. 40.

Mislove, A., S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, J. N. Rosenquist (2011) Understanding the demographics of twitter users., *ICWSM*, 11, p. 5th.

Montoyo, A., P. MartíNez-Barco, A. Balahur (2012) Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments.

Murphy, K. P., et al. (2006) Naive bayes classifiers, *University of British Columbia*, 18.

Nabil, M., M. Aly, A. Atiya (2015) Astd: Arabic sentiment tweets dataset, in: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2515–2519.

Nakov, P., A. Ritter, S. Rosenthal, F. Sebastiani, V. Stoyanov (2016) Semeval-2016 task 4: Sentiment analysis in twitter, in: *Proceedings of the 10th international workshop on semantic evaluation (semeval-2016)*, pp. 1–18.

Nanni, L., S. Ghidoni, S. Brahnam (2017) Handcrafted vs. non-handcrafted features for computer vision classification, *Pattern Recognition*, 71, pp. 158–172.

Noulas, A., S. Scellato, N. Lathia, C. Mascolo (2012) A random walk around the city: New venue recommendation in location-based social networks, in: *Privacy, security, risk and trust (PASSAT), 2012 international conference on and 2012 international confernece on social computing (socialcom)*, Ieee, pp. 144–153.

Ortigosa, A., J. M. Martín, R. M. Carro (2014) Sentiment analysis in facebook and its application to e-learning, *Computers in human behavior*, 31, pp. 527–541.

Ou, Q. (2011) Fusing heterogeneous traffic data: Parsimonious approaches using data-data consistency.

Pandey, A. C., D. S. Rajpoot, M. Saraswat (2017) Twitter sentiment analysis using hybrid cuckoo search method, *Information Processing & Management*, 53(4), pp. 764–779.

Pang, B., L. Lee, S. Vaithyanathan (2002) Thumbs up?: sentiment classification using machine learning techniques, in: *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, Association for Computational Linguistics, pp. 79–86.

Pang, B., L. Lee, et al. (2008) Opinion mining and sentiment analysis, *Foundations and Trends® in Information Retrieval*, 2(1–2), pp. 1–135.

Panter, J. R., A. Jones (2010) Attitudes and the environment as determinants of active travel in adults: what do and don't we know?, *Journal of Physical Activity and Health*, 7(4), pp. 551–561.

Paule, J. D. G., Y. Sun, P. V. Thakuriah (2019) Beyond geotagged tweets: Exploring the geolocalisation of tweets for transportation applications, in: *Transportation Analytics in the Era of Big Data*, Springer, pp. 1–21.

Peersman, C., W. Daelemans, L. Van Vaerenbergh (2011) Predicting age and gender in online social networks, in: *Proceedings of the 3rd international workshop on Search and mining user-generated contents*, ACM, pp. 37–44.

Pfitzner, R., A. Garas, F. Schweitzer (2012) Emotional divergence influences information spreading in twitter., *ICWSM*, 12, pp. 2–5.

Plerou, V., P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, H. E. Stanley (1999) Universal and nonuniversal properties of cross correlations in financial time series, *Physical Review Letters*, 83(7), p. 1471.

Polus, A., J. L. Schofer, A. Ushpiz (1983) Pedestrian flow and level of service, *Journal of transportation engineering*, 109(1), pp. 46–56.

Poria, S., E. Cambria, N. Howard, G.-B. Huang, A. Hussain (2016a) Fusing audio, visual and textual clues for sentiment analysis from multimodal content, *Neurocomputing*, 174, pp. 50–59.

Poria, S., I. Chaturvedi, E. Cambria, A. Hussain (2016b) Convolutional mkl based multimodal emotion recognition and sentiment analysis, in: *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, IEEE, pp. 439–448.

Powers, D. M. (2011) Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation.

Pramanik, S., R. Haldar, A. Kumar, S. Pathak, B. Mitra (2019) Deep learning driven venue recommender for event-based social networks, *IEEE Transactions on Knowledge and Data Engineering*.

Pratiwi, A. R., S. Zhao, X. Mi (2015) Quantifying the relationship between visitor satisfaction and perceived accessibility to pedestrian spaces on festival days, *Frontiers of Architectural Research*, 4(4), pp. 285–295.

Psyllidis, A., A. Bozzon, S. Bocconi, C. T. Bolivar (2015a) A platform for urban analytics and semantic data integration in city planning, in: *International conference on computer-aided architectural design futures*, Springer, pp. 21–36.

Psyllidis, A., A. Bozzon, S. Bocconi, C. Titos Bolivar (2015b) Harnessing heterogeneous social data to explore, monitor, and visualize urban dynamics, in: *Planning Support Systems and Smart Cities: Proceedings of the 14th international conference on computers in urban planning and urban management (CUPUM), CUPUM 2015 conference Cambridge (USA), July 7-10, 2015,, 1-22.(2015)*, MIT.

Quercia, D., L. M. Aiello, R. Schifanella, A. Davies (2015a) The digital life of walkable streets, in: *Proceedings of the 24th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee, pp. 875–884.

Quercia, D., R. Schifanella, L. M. Aiello, K. McLean (2015b) Smelly maps: the digital life of urban smellscapes, AAAI.

Rashidi, T. H., A. Abbasi, M. Maghrebi, S. Hasan, T. S. Waller (2017) Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges, *Transportation Research Part C: Emerging Technologies*, 75, pp. 197–211.

Ravi, K., V. Ravi (2015) A survey on opinion mining and sentiment analysis: tasks, approaches and applications, *Knowledge-Based Systems*, 89, pp. 14–46.

Redmon, J., S. Divvala, R. Girshick, A. Farhadi (2016) You only look once: Unified, real-time object detection, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788.

Redmon, J., A. Farhadi (2017) Yolo9000: better, faster, stronger, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271.

Ribeiro, F. N., L. Henrique, F. Benevenuto, A. Chakraborty, J. Kulshrestha, M. Babaei, K. P. Gummadi (2018) Media bias monitor: Quantifying biases of social media news outlets at large-scale, in: *Twelfth International AAAI Conference on Web and Social Media*.

Rietveld, P., V. Daniel (2004) Determinants of bicycle use: do municipal policies matter?, *Transportation Research Part A: Policy and Practice*, 38(7), pp. 531–550.

Rodríguez, D. A., E. M. Brisson, N. Estupiñán (2009) The relationship between segment-level built environment attributes and pedestrian activity around bogota's brt stations, *Transportation research part D: transport and environment*, 14(7), pp. 470–478.

Roy, K. C., M. Cebrian, S. Hasan (2019) Quantifying human mobility resilience to extreme events using geo-located social media data, *EPJ Data Science*, 8(1), p. 18.

Rui, H., Y. Liu, A. Whinston (2013) Whose and what chatter matters? the effect of tweets on movie sales, *Decision Support Systems*, 55(4), pp. 863–870.

Ryan, D., S. Denman, C. Fookes, S. Sridharan (2009) Crowd counting using multiple local features, in: *Digital Image Computing: Techniques and Applications, 2009. DICTA'09.*, IEEE, pp. 81–88.

Ryan, D., S. Denman, S. Sridharan, C. Fookes (2015) An evaluation of crowd counting methods, features and regression models, *Computer Vision and Image Understanding*, 130, pp. 1–17.

Saif, H., Y. He, H. Alani (2012) Semantic sentiment analysis of twitter, in: *International semantic web conference*, Springer, pp. 508–524.

Saleh, S. A. M., S. A. Suandi, H. Ibrahim (2015) Recent survey on crowd density estimation and counting for visual surveillance, *Engineering Applications of Artificial Intelligence*, 41, pp. 103–114.

Schenk, E., C. Guittard, et al. (2009) Crowdsourcing: What can be outsourced to the crowd, and why, in: *Workshop on open source innovation, Strasbourg, France*, vol. 72, Citeseer, p. 3.

Schwartz, H. A., J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman, et al. (2013) Personality, gender, and age in the language of social media: The open-vocabulary approach, *PloS one*, 8(9), p. e73791.

Sindagi, V. A., V. M. Patel (2017) Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting, in: *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, pp. 1–6.

Sindagi, V. A., V. M. Patel (2018) A survey of recent advances in cnn-based single image crowd counting and density estimation, *Pattern Recognition Letters*, 107, pp. 3–16.

Still, G. K. (2000) *Crowd dynamics*, Ph.D. thesis, University of Warwick.

Still, G. K. (2014) *Introduction to crowd science*, CRC Press.

Thelwall, M. (2013) Heart and soul: Sentiment strength detection in the social web with sentistrength, 2013, *Cyberemotions: Collective emotions in cyberspace*.

Thelwall, M., K. Buckley, G. Paltoglou (2011) Sentiment in twitter events, *Journal of the American Society for Information Science and Technology*, 62(2), pp. 406–418.

Thelwall, M., K. Buckley, G. Paltoglou (2012) Sentiment strength detection for the social web, *Journal of the American Society for Information Science and Technology*, 63(1), pp. 163–173.

Thelwall, M., K. Buckley, G. Paltoglou, D. Cai, A. Kappas (2010) Sentiment strength detection in short informal text, *Journal of the American Society for Information Science and Technology*, 61(12), pp. 2544–2558.

Titos Bolivar, C. (2014) City usage analysis using social media.

Tripathy, A., A. Agrawal, S. K. Rath (2016) Classification of sentiment reviews using n-gram machine learning approach, *Expert Systems with Applications*, 57, pp. 117–126.

Tubbs, J., B. Meacham (2007) *Egress design solutions: A guide to evacuation and crowd management planning*, John Wiley & Sons.

Tyshchuk, Y., W. A. Wallace (2018) Modeling human behavior on social media in response to significant events, *IEEE Transactions on Computational Social Systems*, 5(2), pp. 444–457.

van der Waerden, P., A. Borgers, H. Timmermans (1998) The impact of the parking situation in shopping centres on store choice behaviour, *GeoJournal*, 45(4), pp. 309–315.

Walach, E., L. Wolf (2016) Learning to count with cnn boosting, in: *European conference on computer vision*, Springer, pp. 660–676.

Wang, B., M. Ye, X. Li, F. Zhao, J. Ding (2012a) Abnormal crowd behavior detection using high-frequency and spatio-temporal features, *Machine Vision and Applications*, 23(3), pp. 501–511.

Wang, S.-Y., W.-S. Liao, L.-C. Hsieh, Y.-Y. Chen, W. H. Hsu (2012b) Learning by expansion: Exploiting social media for image classification with few training examples, *Neurocomputing*, 95, pp. 117–125.

Wang, Y., G. H. de Almeida Correia, B. van Arem, H. H. Timmermans (2018) Understanding travellers' preferences for different types of trip destination based on mobile internet usage data, *Transportation Research Part C: Emerging Technologies*, 90, pp. 247–259.

Wegener, M. (2004) Overview of land use transport models, in: *Handbook of transport geography and spatial systems*, Emerald Group Publishing Limited, pp. 127–146.

Weppner, J., P. Lukowicz (2013) Bluetooth based collaborative crowd density estimation with mobile phones, in: *Pervasive Computing and Communications, IEEE International Conference on*, IEEE, pp. 193–200.

Wirz, M., T. Franke, D. Roggen, E. Mitleton-Kelly, P. Lukowicz, G. Tröster (2012) Inferring crowd conditions from pedestrians' location traces for real-time crowd monitoring during city-scale mass gatherings, in: *Enabling Technologies: Infrastructure for Collaborative Enterprises, IEEE 21st International Workshop on*, IEEE, pp. 367–372.

Yang, C., M. Xiao, X. Ding, W. Tian, Y. Zhai, J. Chen, L. Liu, X. Ye (2019a) Exploring human mobility patterns using geo-tagged social media data at the group level, *Journal of Spatial Science*, 64(2), pp. 221–238.

Yang, J., C. Hauff, G.-J. Houben, C. T. Bolivar (2016) Diversity in urban social media analytics, in: *International Conference on Web Engineering*, Springer, pp. 335–353.

Yang, Y., A. Heppenstall, A. Turner, A. Comber (2019b) Who, where, why and when? using smart card and social media data to understand urban mobility, *ISPRS International Journal of Geo-Information*, 8(6), p. 271.

Young, T., M. Palta, J. Dempsey, J. Skatrud, S. Weber, S. Badr (1993) The occurrence of sleep-disordered breathing among middle-aged adults, *New England Journal of Medicine*, 328(17), pp. 1230–1235.

Yu, Y., W. Duan, Q. Cao (2013) The impact of social and conventional media on firm equity value: A sentiment analysis approach, *Decision Support Systems*, 55(4), pp. 919–926.

Yuan, Y. (2014) Crowd monitoring using mobile phones, in: *2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 1, IEEE, pp. 261–264.

Yuan, Y., W. Daamen, D. Duives, S. Hoogendoorn (2016) Comparison of three algorithms for real-time pedestrian state estimation-supporting a monitoring dashboard for large-scale events, in: *Intelligent Transportation Systems, IEEE 19th International Conference on*, IEEE, pp. 2601–2606.

Zahran, S., S. D. Brody, P. Maghelal, A. Prelog, M. Lacy (2008) Cycling and walking: Explaining the spatial distribution of healthy modes of transportation in the united states, *Transportation research part D: transport and environment*, 13(7), pp. 462–470.

Zhan, B., P. Remagnino, S. Velastin, F. Bremond, M. Thonnat (2006) Matching gradient descriptors with topological constraints to characterise the crowd dynamics.

Zhang, C., K. Kang, H. Li, X. Wang, R. Xie, X. Yang (2016a) Data-driven crowd understanding: A baseline for a large-scale crowd dataset, *IEEE Transactions on Multimedia*, 18(6), pp. 1048–1061.

Zhang, C., H. Li, X. Wang, X. Yang (2015) Cross-scene crowd counting via deep convolutional neural networks, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 833–841.

Zhang, J.-D., C.-Y. Chow (2015) Core: Exploiting the personalized influence of two-dimensional geographic coordinates for location recommendations, *Information Sciences*, 293, pp. 163–181.

Zhang, K., Y. Xie, Y. Yang, A. Sun, H. Liu, A. Choudhary (2014) Incorporating conditional random fields and active learning to improve sentiment identification, *Neural Networks*, 58, pp. 60–67.

Zhang, Y., D. Zhou, S. Chen, S. Gao, Y. Ma (2016b) Single-image crowd counting via multi-column convolutional neural network, in: *Proceedings* *of the IEEE conference on computer vision and pattern recognition*, pp. 589–597.

Zhou, E., Z. Cao, Q. Yin (2015) Naive-deep face recognition: Touching the limit of lfw benchmark or not?, *arXiv preprint arXiv:1501.04690*.

Zomer, L. B., W. Daamen, S. Meijer, S. P. Hoogendoorn (2015) Managing crowds: The possibilities and limitations of crowd information during urban mass events, in: *Planning Support Systems and Smart Cities*, Springer, pp. 77–97.

Zuurbier, E. (2019) Understanding pedestrians' perception of crowdedness at mass events: A simultaneous survey and monitoring study into personal, trip and event characteristics.

Summary

Events are getting more popular and more frequent in cities around the world. In the Netherlands in 2017, the number of festivals grew to almost 1000 [5]. These events take place in large areas of the city, they have a common topic, they include sub-events (activities), and they have start and end times and lasts from one day to several days. Examples of events are the national holidays, Soul Live Festival and trade exhibitions. City events can easily attract a large number of people. Event stakeholders, such as the event organizers, police, municipalities and other authorities, and crowd managers are concerned with guaranteeing the safety, comfort and general well being of the attendees. To this end, they enforce predefined crowd management measures that are adaptive to the current state of the event environment and of the participating crowd. This state is measured through information about the factors influencing event planning (Li, 2019) and pedestrian behaviour (Still, 2000; Tubbs & Meacham, 2007; Abbott & Geddie, 2000; Zomer et al., 2015) for crowd management, such as crowd size, density, mobility, emotion, visitor profile, and location. Conventionally, this information is derived from data provided by stewards (operating on the ground during the event) and sometimes pre-installed sensing infrastructures, such as counting systems, Bluetooth/ Wi-Fi sensors, and video cameras. While effective, these solutions suffer from several issues: they provide little information about sentiments, gender and age distribution, they are expensive, they cannot provide Spatio-temporal information, and they are complex to install and maintain.

The advent of web-enabled technologies provides new sources of social media data (e.g. Twitter, Instagram, etc.) that potentially can be used from crowd management planning and operations. Compared with traditional data

[5] https://www.eventbranche.nl/nieuws/aantal-festivals-groeit-tot-bijna-1000-per-jaar-aantal-bezoeken-daalt-miniem-16483.html

sources, social media data are created by the people themselves and are often enriched with Spatio-temporal annotations. They possibly contain rich semantic descriptions without extra sensing infrastructure needed. Such features make social media data a promising source of information for deriving characteristics of the crowd in city events. At the same time, social media data also suffers from issues of sparsity, bias and low accuracy for the geo-locations. How and even if social media data can be used for crowd management purposes is still an open research question.

In this thesis, we hypothesise that social media data can enrich considerably the information needed for crowd management, both in the planning phase and in the operations phase. In this thesis, we investigate to what extent social media are able to provide useful information of crowds for crowd management in city events in terms of four topics, namely event characterisation, attendees' sentiments, crowd size and crowd density estimation.

For the first topic, we aim to characterise city events using social media data in terms of various aspects, including demographics of the crowd (age and gender), visiting motive (city-role, e.g. inhabitants, local or foreign tourists), Spatio-temporal distribution of crowds, and linguistic patterns. As a case study, we use two city events that took place in Amsterdam, namely Sail 2015 and King's Day 2016; we processed data from Twitter and Instagram, and we analyze the derived information with data from the events program, and quantitative data from the crowd management infrastructure. The analysis provides evidence showing that the derived information in terms of age, city-role, post position, PoI's preferences and word use, is consistent with expected results according to the events program and event organizers. However, gender composition, referring to the proportion of male and female, derived from social media data is not in line with expectation according to event programs and organizers. At the same time, the derived information about crowds is influenced by event properties, attendees profile, and place characteristics.

The second topic deals with the sentiments in the crowd. Sentiments are defined by a mental feeling or emotion (Cambria et al., 2017), which can be positive, negative or neutral. Having the word use characteristics analysed, to further mining the semantic characteristics about people in the crowd, such as emotions of individuals, we explore how a social media dataset can be used to estimate the sentiments of people in city events. To this end, we construct a sentiment annotated social media dataset which is collected from a set of events, and use the dataset to examine the effectiveness of state-of-art-methods for sentiment analysis. Our findings show that the machine learning based method Linear Support Vector Classifier (LinearSVC), trained with the constructed sentiment dataset, correctly identified sentiment in 82.3% of social media posts and outperforms other methods in estimating the sentiment of the crowd in city events. We conclude that social media data can be used for sentiment analysis using machine learning techniques. These results are encouraging, and suggest that social media data could be used to help crowd managers better understand crowds attending city events.

The third topic pertains to estimating the number of people in social media images. We construct a social media dataset, compare the effectiveness of face recognition, object recognition, and cascaded methods for crowd size estimation, and investigate the impact of image characteristics on the performance of selected methods. The dataset is constructed considering diverse events fingerprints. The crowd size and image characteristics are annotated for each social media image in the dataset. Results show that object recognition based methods Darknet Yolo reach the highest accuracy (72%) in crowd size level estimation and the highest accuracy (38%) in estimating the specific number of people when less than twenty people appear in an image using social media data in city events. We also found that face recognition and object recognition methods are more suitable to estimate the crowd size for social media images which are taken in parallel view (i.e. the camera and the people being photographed are at a similar height), with full face selfies and in which all the people in the background have the same distance to the camera. However, so-called Cascaded methods are more suitable for images taken from top view with gatherings distributed in gradient. The created social media dataset is essential for selecting image characteristics and evaluating the accuracy of people counting methods in an urban event context. It is also valuable for estimating the density of people in city events using social media.

The final topic is about the estimation of density (as a function of time). Using pedestrian traffic flow theory, we proposed a new method to estimate the density of people using social media data in city events. The approach consists of three estimation strategies, i.e. geo-based, speed-based and flow-based density estimation strategies, that were validated using social media data collected during two city events, i.e. Sail 2015 and Kingsday 2016. Findings show that the flow-based method achieves promising performance in all terrains, both in terms of relative mean difference (from 20% to 120% improvement with regard to other methods) and temporal correlation (between 0.54 and 0.87). At the same time, several factors play a significant role in terms of estimation accuracy and temporal correlation, such as the properties of the terrain, demographics, role and Points of Interest preferences of the crowd. When across all terrains, the maximal performance (i.e. higher temporal correlation and estimation precision than other terrains) is achieved with an equal proportion of male and female in the crowd. The sensitivity analysis showed that the speed-based and flow-based methods achieve optimal performance when including active users at walking distance, and within 30-40 minutes from the temporal windows of observation. According to the findings, the flow-based method is useful to estimate the density of people in the crowd in city events for crowd management.

To summarize, this research provides insights into data, methods, and influencing factors (such as visitor profile, crowd location, mobility, emotion and density) when deriving information about crowds using social media in city events. A new method to estimate the density of people in the crowd using social media has been proposed, validated and verified.

The scientific contributions in this work consist of three categories, i.e. methodological contributions, novel insights and constructed datasets. Firstly, the methodological scientific contribution is a new method to estimate densities from social media data developed by the application of (pedestrian) traffic flow theory concepts. Secondly, the social media data analyses are performed as case studies in this research (Chapter 2 and 5). The data analyses provide a better understanding of how social media data can be used to derive information about crowds. Lastly, a set of social media datasets are collected and annotated in this research (Chapter 2 to 5), from diverse city events. These datasets can be further used for, for instance, developing and verifying new models, studying and analysing cases.

The findings reported in this work have important implications for practice. According to the findings, we recommend crowd managers to include web-based data sources, such as social media data, in crowd management. At the same time, crowd managers should be aware of the biases introduced by social media sources and/or content analysis techniques. To stress the sparsity of social media data deriving information about crowds for crowd management, it is recommended to promote social media usage in city events.

When it comes to crowd management applications, in the event planning phase, information about crowds derived from social media collected in history events can be used for crowd behaviour simulation, inferring crowd management guidelines, and preparing what-if scenarios. The sparsity of geo-referenced social media data which may affect deriving information about crowds can be eliminated by scaling up the number of geo-referenced social media data based on probabilistic strategies discussed in Chapter 5. The bias of social media usage in city events can be reduced (Culotta, 2014; Kuru & Pasek, 2016; Firth, 1993) based on the insights of case studies. In the operational phase, the derived (near) current information about crowds using social media data can be used to monitor crowds, predict accidents, and apply feasible predefined measures.

In future work, we plan to study and compare the effectiveness of investigated methods in various contexts of city events, having different nature, size, and location in the city. We are also going to explore methods to derive more aspects of information from social media about people, such as economic status, occupation, interest topics and social networks. In addition, to deal with social media data sparsity in particular after the growing movement called #noGeo, we plan to apply geo-localization techniques to increase the amount of geo-referenced posts to counter the sparsity of geo-referenced data. Finally, we plan to investigate how social media data can be used in real-time for crowd management.

Samenvatting

Social media gebruiken om menigte te karakteriseren in stadsevenementen voor crowd management

Evenementen worden steeds populairder en vaker gehouden in steden over de hele wereld. In Nederland groeide het aantal festivals in 2017 tot bijna 1000 [6]. Deze evenementen vinden vaak plaats in grote delen van de stad, hebben een gemeenschappelijk onderwerp en bieden verschillende activiteiten aan. De evenementen zijn tijdsgeboden en duren meestal één of meer dagen.

Voorbeelden van evenementen zijn de festivals tijdens de nationale feestdagen, Soul Live Festival en vakbeurzen. Stadsevenementen trekken over het algemeen een groot publiek. Belanghebbenden van evenementen, zoals de organisatoren van evenementen, politie, gemeenten en andere autoriteiten, en menigte handhavers zijn daarom verantwoordelijke om de veiligheid, het comfort en het algemene welzijn van de aanwezigen te waarborgen. Hiervoor worden vooraf richtlijnen opgesteld en maatregelen getroffen om met de massa menigte om te gaan. Dit wordt ook wel crowd management genoemd.

Voor het opstellen van de richtlijnen en veiligheidsmaatregelen voor crowd management wordt er rekening gehouden met verschillende factoren die invloed hebben op de evenementenplanning (Li, 2019) en voetgangersgedrag (Still, 2000; Tubbs & Meacham, 2007; Abbott & Geddie, 2000; Zomer et al., 2015). De richtlijnen en maatregelen dienen wel flexibel te zijn zodat organisatoren en handhavers zich kunnen aanpassen aan de situatie in geval er

[6]https://www.eventbranche.nl/nieuws/aantal-festivals-groeit-tot-bijna-1000-per-jaar-aantal-bezoeken-daalt-miniem-16483.html

veranderingen optreden. Enkele voorbeelden factoren zijn: publieksgrootte, dichtheid, mobiliteit, emotie, bezoekersprofiel en locatie. Deze informatie is vaak beschikbaar door fysieke waarnemingen op het evenement zelf of door gebruik te maken van vooraf geïnstalleerde detectie-infrastructuren, zoals telsystemen, Bluetooth / Wi-Fi-sensoren en videocamera's.

Hoewel deze technieken effectief zijn, is de beschikbare data gelimiteerd. Er is bijvoorbeeld weinig informatie beschikbaar over de gevoelens, het geslacht en de leeftijdsindeling van de aanwezige menigte. Daarnaast zijn deze technieken duur in aanschaf en bieden weinig tot geen informatie aan over tijd en locatie. Bovendien zijn deze technieken complex om te installeren en te onderhouden.

De opkomst van web-enabled technologieën biedt nieuwe mogelijkheden aan om relevante data te verzamelen afkomstig van sociale-media platformen zoals Twitter and Instagram, die gebruikt kan worden tijdens de planning en uitvoering van crowd management. Vergeleken met traditionele databronnen wordt socialmedia data door de mensen zelf gecreëerd en vaak verrijkt met ruimtelijk-temporele annotaties. Ze bevatten mogelijk rijke semantische beschrijvingen zonder dat er extra detectie-infrastructuur nodig is. Dergelijke functies maken sociale-mediagegevens een veelbelovende informatiebron voor het afleiden van kenmerken van de menigte bij stadsevenementen. Maar aan de andere kant kampen sociale-media gegevens ook met problemen zoals schaarsheid, vertekening en lage nauwkeurigheid voor de geolocaties. De vraag hoe en of social media gegevens gebruikt kan worden is nog een open onderzoeksvraag.

In dit proefschrift veronderstellen we dat social media gegevens, in zowel de operationele fase als in de planningsfase, de informatie die nodig is voor crowd management aanzienlijk kan verrijken. We onderzoeken in hoeverre sociale media gegevens gebruikt kan worden om nuttige informatie te verschaffen over menigten voor crowd management bij stadsevenementen waarbij we focussen we op vier onderwerpen: evenementkarakterisering, gevoelens van bezoekers, omvang van het publiek en de inschatting van de massadichtheid.

Voor het eerste onderwerp analyseren we de karakteristieken van stadevenementen, waaronder demografische gegevens van de menigte (leeftijd en geslacht), reden van bezoek, de verdeling van menigten en taalkundige patronen. Voor dit onderzoek gebruiken we de social media gegevens afkomstig van Twitter en Instagram van twee stadsevenementen die plaats hebben gevonden in Amsterdam: Sail 2015 en Koningsdag 2016. De onderzoeksresultaten tonen aan dat karakteristieken zoals leeftijd, rol van de stad, post-positie, PoI's voorkeuren en woordgebruik consistent is met de verwachte resultaten volgens het evenementenprogramma en de organisatoren van evenementen. Daar en tegen komt de geslachtssamenstelling, refererend naar het aandeel van mannen en vrouwen, afgeleid van gegevens van sociale media, niet in overeen met de verwachting volgens evenementenprogramma's en organisatoren. Daarnaast zien we ook dat afgeleide informatie over drukte beïnvloed wordt door evenementeigenschappen, het profiel van de deelnemers en locatie kenmerken.

In het tweede onderwerp is gerelateerd aan de gevoelens van de aanwezige menigte. Sentimenten worden bepaald door een mentaal gevoel of emotie (Cambria et al., 2017), die positief, negatief of neutraal kan zijn. We hebben eerst de karakteristieken van de woordkeuzes geanalyseerd. Om de semantische kenmerken van mensen in de menigte, zoals emoties van individuen, verder te af te leiden, onderzoeken we hoe een dataset bestaande sociale media gegevens kan worden gebruikt om de gevoelens van mensen in stadsevenementen in te schatten. We hebben een sentiment-geannoteerde dataset van sociale media geconstrueerd en gebruiken deze dataset om de effectiviteit van state-of-art-methoden voor sentimentanalyse verder te onderzoeken. Onze bevindingen tonen aan dat de op machine- learning gebaseerde methode Linear Support Vector Classifier (LinearSVC), getraind met de opgebouwde sentimentdataset, het sentiment correct identificeerde in 82,3% van de berichten op sociale media en beter presteert dan andere inschattingsmethoden van het sentiment van de menigte bij stadsevenementen. Hieruit concluderen we dat sociale mediagegevens gebruikt kan worden voor sentimentanalyse met behulp van machine learning-technieken. Deze resultaten zijn bemoedigend en suggereren dat het gebruik van sociale-mediagegevens een positieve bijdrage kunnen leveren om crowdmanagers te helpen bij het beter begrijpen van menigten die naar evenementen in de stad gaan.

Het derde onderwerp heeft betrekking op het inschatten van het aantal mensen op afbeeldingen van sociale media gegevens. We construeren een sociale media dataset, vergelijken de effectiviteit van gezichtsherkenning, objectherkenning en trapsgewijze methoden voor het inschatten van de menigte en onderzoeken de impact van beeldkenmerken op de prestaties van geselecteerde methoden. De dataset is samengesteld uit social media gegevens van verschillende stadsevenementen. De omvang van de menigte en beeldkenmerken worden geannoteerd voor elk social media-beeld in de dataset. De resultaten tonen aan dat Darknet Yolo, gebaseerd op objectherkenning, de hoogste nauwkeurigheid (72%) bereikt bij het schatten van de omvang van het publiek en de hoogste nauwkeurigheid (38%) bij het inschatten van het specifieke aantal mensen wanneer minder dan twintig mensen in een afbeelding verschijnen.

Tevens tonen de resultaten aan dat methoden voor gezichtsherkenning en objectherkenning geschikter zijn om de omvang van het publiek in te schatten voor afbeeldingen die genomen zijn in landscapemode (d.w.z. de camera en de mensen die worden gefotografeerd bevinden zich op een vergelijkbare hoogte), zelfportretten met het gehele gezicht erop en afbeeldingen waarbij alle mensen op de achtergrond dezelfde afstand tot de camera hebben. Daar en tegen zijn trapsgewijze-methoden geschikter voor afbeeldingen die van bovenaf zijn genomen waarbij het publiek verspreid staat. De gecreëerde sociale media dataset is essentieel voor het selecteren van beeldkenmerken en het evalueren van de nauwkeurigheid van methoden voor het tellen van mensen in een stedelijke evenementcontext. Het is ook waardevol voor het inschatten van de dichtheid van mensen bij stadsevenementen.

Het laatste onderwerp gaat over het inschatten van dichtheid (als functie van tijd). Met behulp van de verkeersstroomtheorie voor voetgangers hebben we een nieuwe methode voorgesteld om de dichtheid in te schatten van mensen die sociale-mediagegevens gebruiken bij stadsevenementen. De aanpak bestaat uit drie inschattingsstrategieën: geo-gebaseerde strategieën, op snelheid gebaseerde strategieën en op stroming gebaseerde dichtheidsschattingsstrategieën. Deze strategieën zijn gevalideerd met behulp van socialemediagegevens die zijn verzameld tijdens de twee stadsevenementen Sail 2015 en Koningsdag 2016. De resultaten tonen aan dat de flow gebaseerde methode veelbelovende prestaties behaalt op alle terreinen, zowel het relatief gemiddeld verschil (van 20% tot 120% verbetering ten opzichte van andere methoden) als de temporele correlatie (tussen 0,54 en 0,87). Tevens zien we dat verschillende factoren, zoals de eigenschappen van het terrein, demografische gegevens, rol en POI-voorkeuren van de menigte, een belangrijke rol spelen in de schattingsnauwkeurigheid en temporele correlatie. Op alle gebieden wordt de maximale prestatie (d.w.z. hogere temporele correlatie en schattingsprecisie dan andere terreinen) bereikt met een gelijk aandeel mannen en vrouwen in de aanwezige menigte. De gevoeligheidsanalyse toont aan dat de op snelheid gebaseerde methoden en op flow-based methoden optimale prestaties bereiken wanneer actieve gebruikers op loopafstand worden bijgehouden en zich binnen 30-40 minuten van de temporele observatievensters bevinden. De resultaten tonen aan dat de flow-based methode gebruikt kan worden om een nauwkeurig schatting te maken van de dichtheid van mensen in stadsevenementen.

Samenvattend biedt dit onderzoek inzicht in data, methoden en beinvloedende factoren (zoals bezoekersprofiel, publiekslocatie, mobiliteit, emotie en dichtheid) bij het afleiden van informatie over menigten gebruikmakend van sociale media gegevens bij stadsevenementen. Een nieuwe methode om de dichtheid van mensen in de mensenmassa in te schatten met behulp van sociale media is voorgesteld, gevalideerd en geverifieerd.

De wetenschappelijke bijdragen in dit werk bestaat uit drie onderdelen: de methodologische bijdragen, nieuwe inzichten in het gebruik van gegevens uit social media en geconstrueerde datasets voor analyses. De eerste bijdrage is de methodologisch wetenschappelijke bijdrage waarbij een nieuwe methodiek is voorgesteld om dichtheden van menigte in te schatten uit gegevens van social media door gebruik te maken van toepassing van (voetgangers) verkeersstroomtheorie concepten. De tweede bijdrage is het data analyse onderzoek op de casus studies beschreven in (Hoofdstuk 2 en 5). De data analyses geven een beter inzicht in hoe social media gegevens gebruikt kunnen worden om informatie over menigten af te leiden. Ten slotte wordt in dit onderzoek een reeks gegevenssets van sociale media verzameld en geannoteerd (Hoofdstuk 2 naar 5), bestaande uit diverse stadsevenementen. Deze datasets kunnen verder worden gebruikt voor bijvoorbeeld het ontwikkelen en verifiëren van nieuwe modellen, het bestuderen en analyseren van andere onderzoeksvraagstukken.

De bevindingen in dit onderzoek hebben belangrijke implicaties voor de praktijk. Aan de hand van deze onderzoeksresultaten raden we crowdmanagers aan om web gebaseerde gegevensbronnen, zoals socialmedia gegevens, te gebruiken bij crowdmanagement. Tegelijkertijd moeten crowdmanagers bewust zijn van de onzuiverheden die worden veroorzaakt door gegevens voor sociale media en / of analyze technieken. Om de schaarsheid te benadrukken van sociale media gegevens waarmee informatie over menigten kan worden afgeleid voor crowd management, wordt aanbevolen om het gebruik van sociale media bij stadsevenementen te promoten.

Als het gaat om toepassingen voor crowd management, kan in de planningsfase gebruik gemaakt worden van historisch verzamelde social media data voor een simulatie van het gedrag van de menigte, het afleiden van richtlijnen voor crowd management en het voorbereiden van was-als situaties.

De schaarsheid van geo-gerefereerde sociale-mediagegevens die van invloed kunnen zijn op het afleiden van informatie over menigten, kan worden geëlimineerd door het aantal geo-gerefereerde sociale-mediagegevens op te schalen op basis van probabilistische strategieën die worden besproken in hoofdstuk 5. Het vooroordeel van het gebruik van sociale media bij stadsevenementen kan worden verminderd (Culotta, 2014; Kuru & Pasek, 2016; Firth, 1993) door de verkregen inzichten van de casus onderzoeken. In de operationele fase kan de afgeleide information over drukte afkomstig uit social media gebruikt worden om drukte te monitoren, ongevallen te voorspellen en vooraf gedefinieerde maatregelen te treffen.

In toekomstig onderzoek zijn we van plan de effectiviteit van onderzochte methoden te bestuderen en te vergelijken met verschillende contexten van stadsevenementen met verschillende aard, grootte en locatie in de stad. Tevens gaan we methoden onderzoeken om meer aspecten van informatie uit sociale media over mensen af te kunnen leiden, zoals economische status, beroep, interessethema's en sociale netwerken. Daarnaast zijn we van plan om geo-lokalisatie-technieken toe te passen om de hoeveelheid geo-gerefereerde berichten te vergroten en om de schaarsheid van geo-gerefereerde data tegen te gaan en daarbij met name om de schaarsheid van sociale media in het bijzonder na de groeiende beweging genaamd #noGeo aan te pakken. Ten slotte zijn we van plan om te onderzoeken hoe social media-gegevens in realtime gebruikt kan worden bij crowd management.

Summary in Chinese

概述

近年来，大型城市活动在世界各地越来越受欢迎，越来越频繁，比如：国庆庆典，大型音乐会，运动会，贸易展会等等。2017年荷兰的城市庆典活动增长到近1000个[7]。这些大型活动在城市的大片区域举行，它们有一个共同的主题，由很多小型活动组成，并且都有开始和结束时间，有的持续一天，有的持续好几天。城市活动很容易吸引大量的人，活动的利益相关者（组织者、警察、市政当局等）对人群进行管理，以保证参加者的安全、舒适和健康。为此，他们预先制定人群管理的措施，并根据当前的活动环境和人群的信息来执行这些预定措施。其中人群的状态包括：人群规模、密度、流动性、情绪、人群的年龄分布，性别组成，以及所在位置等等。这些状态是通过影响活动规划(Li, 2019)和行人行为(Still, 2000; Tubbs & Meacham, 2007; Abbott & Geddie, 2000; Zomer et al., 2015)的人群管理因素来衡量的。按照惯例，这些信息可以由现场工作人员（活动期间在现场）提供，有时也来自于预装的传感设施，比如计数系统、蓝牙/Wi-Fi传感器，或者视频摄像头。虽然这些解决方案被广泛使用，但也存在一些问题：它们提供的关于情绪、性别和年龄分布的信息很少，价格昂贵，不能提供时间和空间信息，大规模的安装和维护比较困难。

网络技术的出现提供了新的社交媒体数据源，如Twitter（推特，类似微博）、Instagram（照片墙，类似微信朋友圈）等等，这些数据有可能被应用于制定人群管理的计划并实施这些计划。与传统的数据源相比，社交媒体数据是人们自己创造的，通常附有时间和空间信息，并且可能包含丰富的语义信息（比如谈论的话题等等），而不需要额外的途径来获取这些信息。这样的特点使得社交媒体数据成为提取城

[7]https://www.eventbranche.nl/nieuws/aantal-festivals-groeit-tot-bijna-1000-per-jaar-aantal-bezoeken-daalt-miniem-16483.html

市大型活动中人群信息的一个有潜力的数据源。同时，社交媒体数据也存在着一些问题，例如数据有偏差，附带地理信息（即空间信息）的数据偏少，地理定位精度低等。综上所述，能否将社交媒体数据用于人群管理，如果做到，仍是一个开放的研究问题。

在本论文中，我们假设无论是在人群管理的计划阶段还是在运营阶段，社交媒体数据都可以极大地丰富人群管理所需的信息。在本论文中，我们从大型活动的特征、参会者情绪、人群规模和人群密度这四个主题出发，研究社交媒体能够在多大程度上为大型城市活动的人群管理提供有用的人群信息。

对于第一个主题，我们旨在利用社交媒体数据从不同方面描述城市活动的特征，包括人群的人口统计学特征（即年龄和性别）、参观动机（即城市角色，如居民、本地或外国游客）、人群在活动中的时间和位置分布，以及他们的语言特征，作为案例研究。我们以发生在荷兰阿姆斯特丹的两个大型城市活动为例，即Sail 2015（2015年帆船节）和King’s Day 2016（2016年国王节）；我们分析了来自社交媒体Twitter（推特）和Instagram（照片墙）的数据，并将提取的信息与该活动的项目的信息以及一些预先安装的传感设施的定量数据进行分析。证据表明，提取自社交媒体的信息在年龄、城市角色、位置分布、兴趣点（PoI）的喜好，和语言使用等方面，与活动方案和活动组织者提供的预期结果一致。但是，从社交媒体数据中得出的性别构成，即男性和女性的比例，却不符合预期。同时，提取出的人群信息还受到活动属性、参会者特征和活动场所特征的影响。

第二个话题是关于人群中的情绪。情绪是由心理感受（mental feeling）或情感（emotion）来定义的(Cambria et al., 2017)，它可以是积极的、消极的或中性的。在分析了社交媒体数据中的词语使用特征后，为了进一步挖掘人群的语义特征（如个体的情绪），我们研究了如何利用社交媒体数据集来估计城市事件中人们的情绪。为此，我们构建了一个情绪数据集，这个数据集由一组大型城市活动的社交媒体数据组成，并且对每一条数据所表达的情绪进行人工判断和标注。接着，我们使用该数据集来比较当前基于数据的情绪分析方法的有效性。我们的研究结果表明，机器学习的方法线性支持向量分类器（LinearSVC），经过构建的情绪数据集的训练，能够正确识别出不同的情绪。

第三个主题是关于估算社交媒体图像中的人数。我们构建了一个社交媒体数据集，比较了人脸识别、物体识别和级联方法对人群规模估计的有效性，并研究了图像特征对所选方法性能的影响。数据集的构建考虑了多样化的事件指纹。对数据集中的每张社交媒体图像进行人群规模和图像特征的标注。结果显示，基于对象识别的方法Darknet

Yolo在人群规模水平估计方面达到了最高的准确率（72%），在人数少于20人时的图像中估计具体的人数时，准确率在所有测试方法中最高（38%）。我们还发现，人脸识别和物体识别方法更适合估计人群规模，这些社交媒体图像是以平行视角拍摄的（即相机和被拍摄的人的高度相近），并且是全脸自拍，背景中所有的人与相机的距离相同。然而，所谓的级联方法方法更适合应用于从俯视图拍摄的图像，聚集物呈梯度分布。所创建的社交媒体数据集对于选择图像特征和估算城市活动中人员的数量的准确性至关重要。对于利用社交媒体估算城市活动中的人群密度也很有价值。

最后一个主题是关于人群中人的密度的估计。利用人流理论，我们提出了一种新的方法，利用社交媒体数据估计城市事件中的人流密度。该方法包括三种估计策略，即基于地理的、基于速度的和基于流量的人流密度估计策略。利用两个城市活动，即Sail 2015（阿姆斯特丹2015年帆船节）和Kingsday 2016（阿姆斯特丹2016国王节），期间收集的社交媒体数据进行验证。研究结果显示，无论是相对均值差（相对于其他方法从20%提高到120%）还是时间相关性（在0.54和0.87之间），基于流量的方法在所有观测区域中都取得了好的性能。同时，观测区域的属性、人口统计学特点、人群的参观动机和兴趣点偏好等因素在人群密度的估算精度和时间相关性方面起着重要作用。在所有的观测区域中，在人群中男性和女性比例相同的情况下，人群的密度估算达到了最好的效果（即比其他地形有更高的估计精度和时间相关性）。敏感性分析表明，当纳入活跃用户的步行距离时，以及在30-40分钟的时间窗口内，基于速度和基于流量的人群密度估算策略达到了最佳效果。根据研究结果，基于流量的方法对于估计大型城市活动中人群的密度，进行人群管理是非常有用的。

综上所述，本研究对城市活动中使用社交媒体获取人群信息时的数据、方法和影响因素（如游客概况、人群位置、流动性、情绪和密度等）提出了见解。提出了一种利用社交媒体估计人群密度的新方法，并进行了验证。

这项工作的科学贡献包括三类，即方法学贡献、关于大型城市活动的新的见解,和构建的数据集。首先，方法学上的科学贡献是应用（行人）交通流理论概念，开发了一种从社交媒体数据估计人群密度的新方法。其次，社交媒体数据分析在本研究中作为案例研究进行（第2章和第5章）。通过数据分析，可以更好地理解如何利用社交媒体数据来提取人群信息。最后，本研究收集了一组社交媒体数据集，并对其进行了注释（第2章至第5章），这些数据集来自不同的城市活动。这些数据集可进一步用于开发，分析和验证新的模型、研究和案例等。

这项工作的研究结果对实践具有重要意义。根据研究结果，我们建议在人群管理中纳入互联网数据源，如社交媒体数据。同时，人群管理者应该意识到社交媒体数据源和分析方法会引入的数据偏差。为了减少社交媒体数据在提取人群信息时的稀缺性，建议在城市活动中推广社交媒体的应用。

在人群管理的应用方面，包含两个阶段，即城市活动的策划阶段和操作阶段。在策划阶段，从历史活动中收集的社交媒体数据中可以提取出的人群信息，这些信息可以用于人群行为模拟、推断人群管理准则、准备应急预案。附有地理信息的社交媒体数据的稀缺可能会影响人群信息的提取，但这可以通过基于第5章中讨论的概率策略扩大有限的附有地理信息的社交媒体数据量来消除。社交媒体在城市活动中的使用偏差可以通过历史案例的分析结果来减弱(Culotta, 2014; Kuru & Pasek, 2016; Firth, 1993)。在操作阶段，利用社交媒体数据提取出的（接近）当前人群的信息可以用来预测事故，并应用可行的应急预案。

在未来的工作中，我们计划研究不同的方法在不同性质、规模和地点的城市活动中的有效性。我们还将探索怎样从社交媒体中获取更多关于人们的信息的方法，如经济状况、职业、兴趣话题和社交网络。此外，为了应对社交媒体数据的稀缺，尤其是在名为”#noGeo ”的运动愈演愈烈的今天，我们计划应用地理本地化技术来估算社交媒体数据的地理信息，以获取更多的附带地理信息的社交媒体数据。最后，我们计划研究如何实时将社交媒体数据用于人群管理。

About the author

X. Gong was born in Jiajiang, Sichuan, on 20th December 1983. He completed the BSc. of Computer Science at the University of Electronic Science and Technology of China in 2006.

After his graduation, he worked in the ICT industry for about 6 years, where he worked as a software developer and technical sales in Asia and Europe.

In 2013, he came to Delft to study his MSc. in Computer Science at the Delft University of Technology, where he specialised in Information Architecture. He obtained his MSc. degree in April 2016, with the thesis focussed on deriving human activity patterns across cities using social media data.

On 20th June 2016, he started the PhD research in the group of Transport and Planning at the Delft University of Technology, funded by the project of Allegro (ERC: 669792). He submitted the dissertation draft on 24th March 2020.

His research interests include (web) data crawling, social data analysis, and crowd behaviour in city events.

Publications

Journal papers

1. **Gong, Vincent X.**, Jie Yang, Winnie Daamen, Alessandro Bozzon, Serge P. Hoogendoorn, and Geert-Jan Houben. ”Using social media for attendees density estimation in city-scale events.” IEEE Access 6 (2018): 36325-36340.

2. **Gong, Vincent X.**, Winnie Daamen, Alessandro Bozzon, and Serge P. Hoogendoorn. Crowd Characterization Using Social Media Data in City-Scale Events for Crowd Management. Travel Behaviour and Society (2020): 192-212.

3. **Gong, Vincent X.**, Winnie Daamen, Alessandro Bozzon, and Serge P. Hoogendoorn. ”Estimate Sentiment of Crowds from Social Media during City Events.” Transportation Research Record (2019): 0361198 119846461.

4. **Gong, Vincent X.**, Winnie Daamen, Alessandro Bozzon, and Serge P. Hoogendoorn. ”Counting people in the crowd using social media images for crowd management in city events.” Transportation (2020): Under review.

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series, For a complete overview of more than 250 titles see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Gong, X., *Using Social Media to Characterise Crowds in City Events for Crowd Management*, T2020/14, TRAIL Thesis Series, the Netherlands

Rijal, A., *Managing External Temporal Constraints in Manual Warehouses*, T2020/13, September 2020, TRAIL Thesis Series, the Netherlands

Alonso González, M.J., *Demand for Urban Pooled On-Demand Services: Attitudes, preferences and usage*, T2020/12, July 2020, TRAIL Thesis Series, the Netherlands

Alwosheel, A.S.A., *Trustworthy and Explainable Artificial Neural Networks for choice Behaviour Analysis*, T2020/11, July 2020, TRAIL Thesis Series, the Netherlands

Zeng, Q., *A New Composite Indicator of Company Performance Measurement from Economic and Environmental Perspectives for Motor Vehicle Manufacturers*, T2020/10, May 2020, TRAIL Thesis Series, the Netherlands

Mirzaei, M., *Advanced Storage and Retrieval Policies in Automated Warehouses*, T2020/9, April 2020, TRAIL Thesis Series, the Netherlands

Nordhoff, S., *User Acceptance of Automated Vehicles in Public Transport*, T2020/8, April 2020, TRAIL Thesis Series, the Netherlands

Winter, M.K.E., *Providing Public Transport by Self-Driving Vehicles: User preferences, fleet operation, and parking management*, T2020/7, April 2020,

TRAIL Thesis Series, the Netherlands

Mullakkal-Babu, F.A., *Modelling Safety Impacts of Automated Driving Systems in Multi-Lane Traffic*, T2020/6, March 2020, TRAIL Thesis Series, the Netherlands

Krishnakumari, P.K., *Multiscale Pattern Recognition of Transport Network Dynamics and its Applications: A bird's eye view on transport*, T2020/5, February 2020, TRAIL Thesis Series, the Netherlands

Wolbertus, *Evaluating Electric Vehicle Charging Infrastructure Policies*, T20 20/4, February 2020, TRAIL Thesis Series, the Netherlands

Yap, M.D., *Measuring, Predicting and Controlling Disruption Impacts for Urban Public Transport*, T2020/3, February 2020, TRAIL Thesis Series, the Netherlands

Luo, D., *Data-driven Analysis and Modeling of Passenger Flows and Service Networks for Public Transport Systems*, T2020/2, February 2020, TRAIL Thesis Series, the Netherlands

Erp, P.B.C. van, *Relative Flow Data: New opportunities for traffic state estimation*, T2020/1, February 2020, TRAIL Thesis Series, the Netherlands

Zhu, Y., *Passenger-Oriented Timetable Rescheduling in Railway Disruption Management*, T2019/16, December 2019, TRAIL Thesis Series, the Netherlands

Chen, L., *Cooperative Multi-Vessel Systems for Waterborne Transport*, T201 9/15, November 2019, TRAIL Thesis Series, the Netherlands

Kerkman, K.E., *Spatial Dependence in Travel Demand Models: Causes, implications, and solutions*, T2019/14, October 2019, TRAIL Thesis Series, the Netherlands

Liang, X., *Planning and Operation of Automated Taxi Systems*, T2019/13, September 2019, TRAIL Thesis Series, the Netherlands

Ton, D., *Unravelling Mode and Route Choice Behaviour of Active Mode Users*, T2019/12, September 2019, TRAIL Thesis Series, the Netherlands

Shu, Y., *Vessel Route Choice Model and Operational Model Based on Optimal Control*, T2019/11, September 2019, TRAIL Thesis Series, the Nether-

lands

Luan, X., *Traffic Management Optimization of Railway Networks*, T2019/10, July 2019, TRAIL Thesis Series, the Netherlands

Hu, Q., *Container Transport inside the Port Area and to the Hinterland*, T2019/9, July 2019, TRAIL Thesis Series, the Netherlands

Andani, I.G.A., *Toll Roads in Indonesia: transport system, accessibility, spatial and equity impacts*, T2019/8, June 2019, TRAIL Thesis Series, the Netherlands

Ma, W., *Sustainability of Deep Sea Mining Transport Plans*, T2019/7, June 2019, TRAIL Thesis Series, the Netherlands

Alemi, A., *Railway Wheel Defect Identification*, T2019/6, January 2019, TRAIL Thesis Series, the Netherlands

Liao, F., *Consumers, Business Models and Electric Vehicles*, T2019/5, May 2019, TRAIL Thesis Series, the Netherlands

Tamminga, G., *A Novel Design of the Transport Infrastructure for Traffic Simulation Models*, T2019/4, March 2019, TRAIL Thesis Series, the Netherlands

Lin, X., *Controlled Perishable Goods Logistics: Real-time coordination for fresher products*, T2019/3, January 2019, TRAIL Thesis Series, the Netherlands

Dafnomilis, I., *Green Bulk Terminals: A strategic level approach to solid biomass terminal design*, T2019/2, January 2019, TRAIL Thesis Series, the Netherlands

Feng, Fan, *Information Integration and Intelligent Control of Port Logistics System*, T2019/1, January 2019, TRAIL Thesis Series, the Netherlands