Mengmeng Ye

Aspect-Based Review Extraction for E-Commerce Products
Aspect-Based Review Extraction for E-Commerce Products

by

Mengmeng Ye
4407326

in partial fulfilment of the requirements for the degree of

Master of Science in Computer Science / Data Science and Technology

at the Delft University of Technology,
to be defended publicly on Friday 25, Aug, 2017.

Supervisor: Prof. Dr. ir. G.J. Houben WIS, EEMCS, TU Delft
Thesis committee: Dr. C. Lofi WIS, EEMCS, TU Delft
Dr. Marco Zungia ES, EEMCS, TU Delft

Copyright © 2017 by M.Ye
An electronic version of this thesis is available at http://repository.tudelft.nl/.
PREFACE

This is the thesis of Computer Science Master Programme, Data Science & Technology track in Faculty Electrical Engineering, Mathematics & Computer Science at Delft University of Technology. It marks the end of a two years phase of gaining knowledge and personal experiences in the Netherlands. This was an unique opportunity that will be always remembered.

I would like to thank Dr. Christoph Lofi for the guidance and motivation during the thesis project; Dr. Nava Tintarev for helping me with my evaluation design; and all the participants for the time and helping me with the experiment.

I would also like to thank all my friends, my parents and especially my boyfriend Liam Mac an Bhaird for helping me go through all the frustrating time in my study, and supporting me.

Mengmeng Ye
August 2017
SUMMARY

With the development of Internet, more and more people do shopping online. Shopping websites such as Amazon.com allow customers to leave reviews to the products after purchase, which is able to help other people understand the products better and make purchase decision. However, there are often hundreds or thousands of reviews under popular products, and the overview average scores (e.g. 3 stars out of 5) of the products from the reviews do not provide enough information to the users. People can not find information related to different aspects of a product (such as “picture quality” aspect of product “camera”) by just looking at the overview score. If they want to know more they need to read through the reviews. However, by simply reading the top ranked helpful reviews or several random reviews, people will get biased opinions, because their preference may not match with the top voted or average opinions. To understand the products more without bias, the users need to read through a lot of reviews, which is tedious to do. Therefore, it may help the users understand products faster if there is an approach which can extract information about different aspects of the products automatically for the users, and show users the different opinions on the products so that they can decide themselves.

Under this circumstance, this thesis proposes two research questions: is it possible to automatically extract aspects of online products from reviews which are also consistent with the manual modeling of products in this domain? Is it possible to build an aspect-based extraction system to aggregate opinions on online products from reviews, which can help users understand the products faster? To answer these two questions, this thesis introduces a prototype system which is developed to automatically aggregate people’s opinions from reviews on different aspects of online product, in order to help users understand the products faster. The dataset used in the system is movie review data from Amazon.com. Then two evaluations of the system are conducted in order to answer the two research questions.

The prototype system is divided into three parts, aspect extraction, sentiment analysis, and review clustering. In aspect extraction, the system detects which aspect keywords should be used to describe the products based on the reviews of them, and how to detect the aspects mentioned in the reviews. After that, the system performs sentiment analysis to analyze people’s attitudes towards every aspect of the products. After sentiment analysis phase, reviews which share similar opinions are clustered to the same groups to be presented to the users.

In the end, to answer the proposed research questions, the thesis introduces two evaluations: aspect extraction evaluation and system evaluation. In the aspect evaluation, it is divided to two parts. First, to evaluate the performance of the aspect extraction, the aspects found by the prototype system are compared with the aspects manually found by the author. Then, participants with topic modeling knowledge are invited to model the movie products with aspects they think are important, in order to answer whether it is possible to build a system which can automatically extract aspects of the products from reviews and are consistent with human modeling. In the system evaluation, a user case study is conducted to evaluate the system, where each participant is asked to complete two tasks one by one with the help of two different reviews (ranked reviews online and system generated reviews) of two different movies that they didn’t watch before, in order to answer whether the system can help users understand products faster. In the end, the evaluations show that, the prototype system successfully detects most of the aspects appearing in the reviews, which are also consistent with manual modeling. Also, the prototype does help users understand products faster. Even though the prototype system uses dataset from movie reviews, it shows that it has potential to be generalized to products under other categories.
Aspect-based Review Extraction for E-Commerce Products

Mengmeng Ye
Delft University of Technology
M.Ye-1@student.tudelft.nl

ABSTRACT
Nowadays, more and more products are sold online. Under popular products, there are normally hundreds or even thousands of reviews left by the previous customers. These reviews help potential buyers understand the products better and make the purchase decision. However, most shopping websites only give an overview score (e.g. 3 star out of 5) of a product besides the reviews, which does not provide enough information for people to understand different aspects of the product. For example, if the users want to know more about the ‘picture quality’ aspect of a camera product, they need to read the reviews. However, by reading the top ranked reviews or some random reviews, they may get biased information. Professional people may have high demands in ‘picture quality’ while the top ranked reviews may be amateurs who are easy to get satisfied. However, it is tedious for the users to read all of the reviews to get the right information. Moreover, customers tend to have different opinions on one product, but the current shopping websites do not cluster the reviews which share similar opinions together and present it to the users. Instead, they ignore the conflicted opinions in the reviews by simply averaging the scores given by all the reviewers.

Under this circumstance, this thesis proposes to research questions: is it possible to automatically extract aspects of online products which are also consistent with the manual modeling of products in this domain? Is it possible to build a system to aggregate opinions on online products from reviews, which actually can help users understand the products faster? To answer the two research questions, it is necessary to build a prototype system which can extract information and describe different aspects of a product to help users understand the product. This thesis first develops such a system. The system is divided into three steps, aspect detection, sentiment analysis and reviews clustering. In brief, the system first determines which aspects should be used to describe the product, and then calculates people’s sentiment towards these aspects, in the end the system clusters reviews which share similar opinions together. To answer the research questions, two evaluations of the system are conducted in the end, which shows the system has big potential in helping users understanding online shopping products.

KEYWORDS
opinion mining, aspect-based sentiment analysis, opinion clustering

1 INTRODUCTION
With the rapid development of Internet, more and more people are used to do shopping online. Shopping websites such as Amazon.com encourage customers to leave reviews for the products after purchase [1], in order to help other people understand the products better and make purchase decision. However, there are often hundreds or thousands of reviews under a popular product, and the current shopping websites only present an overview score of the product (such as 4 stars out of 5) to users without mentioning the ratings of different aspects of the product. For example, assume that a camera product has only an overview score of 4 star, people can not get information about different aspects of the product (such as “picture quality” and “battery life”) by just looking at the score. To know more about it, they have to read reviews. However, they may not agree with the top ranked reviews or some random reviews. If the top ranked reviews are amateurs who do not care so much about picture quality of the camera, and the user is a professional who cares about it a lot, then the reviews will give him wrong impression about the product. The user needs to read through a lot of reviews to find the right information about the different aspects of a product. As a result, people can not understand the products comprehensively in a short time, and it is impossible for them to read through all of reviews to get an unbiased opinion on the products.

Moreover, traditional ways of opinion mining in reviews tend to ignore the conflicts of opinions appearing in the reviews. Some works rank the strength of reviews and present it to users [28], simply clustering reviews to “negative”, “positive” [16], or calculating the average score for every aspect of the products [22]. However, the quality of some products may be controversial, a user may not agree with the highly ranked or averaged review opinions. As mentioned before, a professional may not agree with the opinion of an amateur. Therefore, these methods are not able to present the conflicted opinions appearing in the reviews to users.

Under this circumstance, this thesis tries to find an approach which can improve the situation by automatically aggregating opinions on online products from reviews and present it to users. The opinions should not be simply classified as “negative” and “positive” opinions. For example, “The camera is good but battery doesn’t last long” indicates that the battery does not have good quality, even though the review is positive towards this product as a whole. Therefore, for people who care more about the battery life of camera, simply clustering this review to “positive” will not be helpful. It will be better if such an approach can cluster similar opinions together based on different aspects of the products (e.g. reviews positive towards “battery life” are clustered together, reviews negative towards “picture quality” are clustered together) rather than based only on “negative” and “positive” of the whole product. After presenting the different opinion groups to the users, they can decide themselves which cluster of reviews they care about most. The thesis proposes the following two research questions:

- RQ1. Is it possible to automatically extract aspects of online products from reviews which are also consistent with the manual modeling of products in this domain?
- RQ2. Is it possible to build an aspect-based extraction system to aggregate opinions on online products from reviews, which indeed can help users understand the products faster?

To answer these two questions, a prototype system is developed and evaluated in this thesis. The prototype system automatically aggregates people’s opinions on different aspects of
online products to describe the products, by using unstructured review meta-data [24], here the aspects should be related to the products, not with the reviewers themselves (e.g. their own life experience related to the movie), nor the service of the websites (such as delivery).

To answer the first research question, the aspects of products generated by the prototype system are evaluated. In the evaluation, participants of a large scale who have background in data modeling are invited to model the corresponding products with aspects they think are important, in order to evaluate if the aspects generated by the prototype system are consistent with the manual modeling. The result shows that the result generated by the prototype system does match with the manual modeling.

To answer the second question, a user case study is conducted. Participants are invited to use both the reviews generated by the prototype system and the top 10 ranked helpful reviews on shopping website of the corresponding products, in order to evaluate if the prototype system can indeed help users understand products faster. The result of the evaluation indicates that such a system has potential in helping users understand products faster, and remember more details of the products.

The prototype system is divided into three parts, the first part is aspect detection, the prototype system decides which aspects should be used to describe the products. For instance, the aspects “picture quality” “battery life” can be used to describe camera product, “acting” “directing” can be used to describe movie product, possibly because they are mostly mentioned and cared by the customers. The second part is sentiment analysis. In brief, sentiment analysis detects what the writer’s attitude is towards a topic in his review (such as positive, negative or calculates a sentiment polarity score) by analyzing the words he uses (e.g. “good” is positive while “bad” is negative). The prototype system detects which aspects appeared in a review and what their sentiments are, then gives ratings of these aspects. For example, “The camera is good but battery doesn’t last long” may indicate that the reviewer has a low score (e.g. around 3.0 or lower) rating towards “battery life”. In the end, based on each review’s aspect sentiment score, the prototype system is able to cluster reviews which share similar opinions together (e.g. one group cares more about picture quality, one groups cares more about the battery). This prototype system uses dataset which covers only reviews in the Movies and TV category on Amazon.com, provided by [24], but the prototype system should be able to be generalized to products in other categories on other shopping websites as well.

This thesis can benefit different kinds of users, such as the website itself and potential customers, because potential customers can understand product faster and thus make faster purchase decision. As well as product managers who are interested in how to improve their products. Basically, by answering the two research questions mentioned before, the thesis gives an insight that such a prototype system may benefit any user who is interested in the opinions behind the online product reviews.

2 RELATED WORK

There are a lot of works related to different opinion mining topics, such as aspect-based sentiment analysis, which analyzes people’s attitudes towards each aspect of the products. As well as text clustering, which clusters texts with similar topics into the same group. However, there is no such a system yet which combines these parts together and works automatically to aggregate people’s opinions on products from reviews, in order to help users understand products faster. Most works only focus on one single part. Therefore, this section discusses the related work of different parts of the prototype system, and research the state-of-the-art techniques which suit the system best.

2.1 Aspect Extraction

As mentioned before, in sentiment analysis of opinion mining field, some works simply classify the sentiment of a whole text document to “positive” or “negative” sentiment. Some of the works manually label documents with “negative” or “positive” and convert the documents to vectors of words as training data, then using supervised learning method [19][26] to classify other documents into the corresponding sentiment group. Some works use unsupervised learning method [13] which does not require manual labeling. However, these methods work in document sentiment analysis of news or blogs, but do not work for reviews. For example, a review classified to positive does not mean this review is positive towards every aspect of the product. “The camera is good but the battery does not last long” indicates that the writer is positive towards the product, but negative towards the aspect “battery life”. Thus, simply labeling this review as “negative” or “positive” leaves out important information.

Therefore, aspect-based sentiment analysis becomes more popular when it comes to analyzing opinions in the reviews. The aspect-based sentiment analysis firstly extracts the aspects of the product mentioned in the reviews (such as “battery life” “picture quality” for camera product, “acting” “directing” for movie product), which will be used to describe the products later. Then the method performs sentiment analysis for each aspect mentioned in the reviews (such as “battery life”: 3 out of 5 star). Therefore, this method seems to meet the needs of this thesis. However, to find the aspects to be used to describe the products, some works use rather simple method by just manually selecting the aspects and the wordset related to the aspects [34], e.g., the authors manually read through the reviews and decide “acting” and “directing” are the aspects mentioned in the reviews of movies, which should be used to describe the movie products. Then the authors list a wordset related to each aspect, such as “actor” “perform” “performance” are words related to the aspect “acting”. In this method, once a word in the reviews matches with any word in an aspect wordset, it is considered this aspect is mentioned in the review. For example, “the performance of the actor is awesome”, in this review there are two words “performance” “actor” which belong to the “acting” aspect wordset, therefore the “acting” aspect is considered mentioned in this review. This method is accurate, but requires a lot of human work and restricts the system to a specific product domain. Some other works use NLP (Natural Language Processing) based method to firstly find all the nouns mentioned in the reviews (since aspects are mostly nouns), then perform frequency analysis to detect the frequently mentioned nouns [23]. For example, “actor” is a noun and it appears frequently in the reviews, then it is considered an aspect. This seems to suit
the system’s need most. Other works use supervised learning based techniques, the authors first manually label words or reviews with different aspects, then classify words into different aspects groups or classify the whole reviews into different aspects groups (aspects are found manually) [15][18], saving a lot of human work. However, this method is normally restricted to a single domain or product category, because the training data is only for products from a single category, e.g. camera product. Then this data can not be used to classify movie reviews.

As mentioned, most of these methods still more or less require human input, either in selecting aspects or aspect related words labeling. Moreover, some of the works have problems to be generalized to products of all categories and will restrict the system to a specific product domain. To answer the research questions, the first goal of this thesis is to develop a completely automatic system which can aggregate people’s opinions on products from reviews. To achieve this, a combination of NLP-based frequency analysis and similarity analysis is used in the thesis to extract the aspects and find the aspect related wordset, which does not require any human input.

2.2 Sentiment Analysis

Sentiment Analysis is a rather broad field, here only the sentiment analysis in opinion mining field is discussed. As mentioned before, sentiment analysis is a process which detects writer’s attitude (e.g. negative, positive or polarity score) towards a topic in his text. Generally, sentiment analysis in opinion mining is divided into two parts, one is supervised learning based analysis, one is dictionary based. In machine learning based field, the works can be divided to supervised learning and unsupervised learning field. In supervised learning field, the researchers normally manually classify the documents to related aspect field, then manually label each aspect of each document with sentiment label (negative, positive) as training data [23]. By training the data using popular machine learning methods, they are able to detect which aspects are mentioned in a document, and what people’s attitudes are towards them. The unsupervised learning methods are used when it is difficult to acquire manually labeled documents such as [37], they use LDA model to extract the topics from reviews automatically, then use the slide window to find the sentiment words around each topic word. However, the problem of LDA and other topic modeling methods is it tends to find the topic of different reviews rather than the aspects of these reviews, i.e., it tends to find topics related to different movies (e.g. “action related movie” “documentary related movie”), rather than their aspects (e.g. “acting” “directing”). In general, the problem of machine learning-based approach is it requires manual labeling, and it is not easy to be generalized to reviews under other product categories, since the aspect related words and sentiment words labeled are domain-specific.

The other approach is dictionary-based approach. Some researchers use existent dictionary with sentiment value such as SentiWordnet [2], which is an expanded version of WordNet. In SentiWordnet, every word is stored with its sentiment scores, for example, for word “breakdown”, it has PosScore=0.0, NegScore=0.25, ObjScore=0.75 (scale 0-1), which means the word “breakdown” has negative score of 0.25, and objective score 0.75. Some works either use the existent dictionary such as SentiWordNet to calculate the sentiment value for each aspect detected [34], or build their own sentiment dictionary [35]. The disadvantage of dictionary-based approach is obvious, i.e., it does not consider the context of the words or the domain. Although their are corpus of different topics (such as finance and news), the thesis intends to build system which can be generalized to all kinds of products, therefore should not use an oriented corpus.

Although the dictionary-based approach has its limitations, comparing to machine learning-based techniques, it is completely automatic and does not restrict the system to a specific domain. To develop an automatic prototype system to answer the research questions, the dictionary-based approach is selected in this thesis.

2.3 Review Clustering

Most sentiment analysis in opinion mining ignores the conflicts of opinions in the reviews, by simply calculating an average or weighted sentiment score of an aspect. Although there are a lot of works related to text clustering, there is hardly a research which applies clustering in the review analysis field. In the text clustering field, a conventional approach is bag-of-words [3], which transforms each document into a vector of words, and is able to calculate the distance between documents based on their vector. For example, if there are only two reviews in the database: “the camera is good” “the camera is bad”, the first review has a vector (the: 1, camera: 1, is: 1, good: 1, bad: 0), the second review has a vector (the: 1, camera: 1, is: 1, good: 0, bad: 1). However this method generates very high dimensional dataset when there are a lot of documents, which suffers from curse of dimensionality [17]. Moreover, it only calculates similarity between the documents, ignoring the sentiment in them. For example, “the story is bad” and “the story is not bad” can be close to each other in the bag-of-words calculation, because they use similar words, but they express actually very different opinions in reviews.

Another approach is feature-based clustering, which selects the feature words or keywords out from the documents and cluster documents with similar feature words into one group [20], efficiently reduce the dimensionality of the dataset, this is similar to what the prototype system needs, because the system selects the aspects of the products from reviews first. However, the problem of it is similar to that of bag-of-words: this approach does not consider the sentiment of the document. Therefore this approach may cluster reviews with completely different opinions to the same group together (“story is bad” and “story is not bad”). Topic modeling based approach is also a popular approach [11], which extracts the potential topics out from each document, each topic is presented as a set of words. Then the model compares the similarity of topics between documents. Similarly, topic modeling does not consider the sentiment of the reviews neither and it tends to find the topic based on different movies rather than the aspects. Also, as mentioned before, there is hardly a research considering clustering reviews with similar opinions together, these works do not focus on people’s opinions towards different aspects of the products.
Therefore, the prototype system chooses to cluster reviews based on the aspect vector of each review, which is similar to the feature-based clustering. This method not only reduces the dimensionality, but also focuses on reviewers’ opinions on different aspects of the products.

3 PROPOSED SYSTEM

As mentioned before, to answer the two research questions, it is necessary to build a prototype system which can aggregate people’s opinions from reviews on different aspects of products, in order to help users understand products more quickly. The system is divided into three parts, aspect detection, sentiment analysis and review clustering. The overview system structure is shown in Figure 1.

3.1 Dataset

The prototype system uses dataset which includes product reviews (ratings, text, helpfulness votes etc.), product information (descriptions, category information etc.) from online shopping website Amazon [24], under category Movie and TV, spanning from May 1996 to July 2014. There are in total 1,697,533 reviews, 50,052 movies in the dataset. The dataset includes only movies with more than 5 reviews and the reviews from users who have given more than 5 reviews. Although it focuses on products of one category, the prototype system should be able to be generalized to products under other categories on other shopping websites as well.

3.2 Aspect Detection

First, the prototype system needs to decide which aspects should be used to describe the products. Therefore, the system analyzes all the reviews under all the products of one category (movie and TV category in this case) to determine which aspects of the products are the most important. Each aspect of the products should not be restrained to a single word. For example, in movie reviews, the review “the acting is awesome” and “the actor plays such a great role” both indicate the “acting” aspect of movie, even the two reviews use different words “actor” and “acting” to describe this aspect. Therefore, an aspect of the products should contain multiple related words in a group (denoted as “wordset” in the following sections). To do it, the system first selects high-ranked aspect words mentioned in the movie reviews as core aspect words (such as “acting”, “directing” etc.), and then finds the words related to each core aspect word to form an aspect wordset (e.g., “perform” “actor” “acting” can be in the wordset of aspect “acting”). A lot of works have been done in the field of keyword extraction and topic modeling [7], these two methods both are able to extract information out of documents, which seem to be similar to what the prototype system needs to do (extract aspect words out of reviews). However, topic modeling fails to extract the topic wordset from the reviews based on different aspects. Instead, it tends to extract topics based on different movies (i.e. action movies, document movies). Therefore, the keywords extraction seems to be helpful, which is chosen by this thesis.

Before extracting the core aspect words, lemmatization is performed on every word in the reviews. Lemmatization helps to find the original form of the word. The reason for not using stemmer is, stemmer stems the words without considering the actual readable form of the word. The stemmed words are normally not recognizable by other natural language processing tools (e.g. “presumably” to “presume”). While lemmatization is able to find the readable form of words (e.g. “presumable” to “presume”), aiming at removing inflectional endings only and to return the base or dictionary form of a word. To extract the core aspect words, the system uses a combination of TF-IDF method [32] and words similarity calculation based on WordNet [25][36]:

In TF-IDF, the importance of a term (word) in a document depends on the frequency of this word appearing in the specific document and the reversed frequency of this term in all the documents. In brief, if a term appears often in one specific document but not often in other documents, then this term is more likely to be important and representative for this specific document, the higher the score is, the more representative a word is.

In the prototype system, to find the core aspect words, all of the movie reviews are considered as one document (movie document), which is where the core aspect words should come from. The other documents to be compared with the movie document come from news articles database provided by scikit-learn [27], the news articles cover 20 topic categories to represent natural language and words used in daily life. The logic behind this method is, if a word appears often in the movie review document but less in the news articles, then it is representative for the movie reviews.
However, this does not necessarily mean a high TF-IDF score word is a core aspect word, the reason is some words are representative for movie reviews but not related to the movie product itself. For example, some people tend to associate their own life experience with the movie when writing reviews: “I’ve always wanted to become a spy in my life after watching this movie when I was young”, words such as “life” is also a highly ranked TF-IDF word. Also sometimes the reviewers mention “package” and “delivery service” which is not related to the product itself. Therefore, the similarity between the words and the product (“movie” in this case) is also taken into account. The similarity calculation [36] is based on WordNet [25]; words are stored as a net in WordNet, this method calculates a similarity score denoting how similar two word senses are, based on the depth of the two senses in the taxonomy and that of their Least Common Subsumer (most specific ancestor node). In brief, if two words are similar two each other, the path between them should be short. For example, the word “acting” is closer to “movie” than the word “life”. There are different methods that are related to word similarity such as association rules [38] and word2vec [10]. However, association rules and word2vec tend to mix up different nouns together, because different aspects normally have similar context. For example, “the movie is amazing” and “the delivery is amazing” have similar neighboring words for the two words “movie” and “delivery”, therefore they are considered similar, but “delivery” is actually not related to the movie product itself and should not be considered an aspect. Other methods are highly relied on human effort, either manually select related words [34] or need manual labels to perform classification [8]. This thesis intends to develop a prototype system which can work completely automatically, therefore the similarity method based on WordNet is chosen. An importance score of the words is simply represented as:

\[ \text{Score} = TF - IDF + \text{similarity} \]  

(1)

In the end, the scores will be calculated for all the words mentioned in the reviews, as shown in Table 1. The top 5 ranked words are chosen as core aspect words for the movie product, which are “acting”, “directing”, “scenery”, “character”, “storyline”. The reason to choose the top 5 words is: the importance scores of the aspect words drops significantly after the top 5. Also, an experiment is conducted which can be found in section 4. 180 participants are asked to model movie with aspects that they think are important in describing movie quality, the top 5 aspects considered important by the participants match with the results found by the system. Moreover, the aspects of movie appearing in the reviews are manually found by the author, the top aspects found by the system match with 5 out of 7 of them, more detailed information will be mentioned in section 4. Therefore, the aspects found by the system are considered indeed descriptive in describing movie, and are consistent with manual modeling.

Now that the prototype system has found the top 5 core aspect words mentioned in the movie reviews, to form 5 aspect wordsets, the system needs to find related words for each aspect. For aspect “acting”, the related words can be “act”, “perform”, etc. The related words are also found by similarity calculation, which is still based on Wu-Palmer Similarity in WordNet. The word which has a similarity over or equals to 0.75 (the scale is 0-1) with core aspect word is considered to be related to the aspect, and thus should be put into this aspect wordset. The reason to choose 0.75 is, when the threshold is lower than 0.75, the amount of related words increase dramatically, therefore 0.75 is selected as an elbow point. It is possible a word can fall in different aspects, same words in different aspects are allowed. The example wordset for aspects are shown in Table 2.

### 3.3 Sentiment Analysis

In the sentiment analysis phase, the prototype system first scans which aspects are mentioned in a review, by matching every word in the review with the aspect wordsets generated in the last phase. If a word appears in the review belongs to an aspect wordset, then this aspect is considered to be mentioned in the review.

To detect the sentiment towards each aspect mentioned in a review, the system tries to find the adjectives and adverbs appeared in the clause (separate by comma) where the aspect wordset word is found, the reason is that the combination of adjectives and adverbs works well in sentiment analysis [5], and they are good indicators of opinions [21]. Therefore a part-of-speech tagging is performed to each word in the review to find its syntactic function (i.e. noun, adjective etc.), by using NLTK toolkit [6]. In brief, the system scans every word in a review to check if there is a match with any aspect wordset. If there is a match, then the system searches for adjectives and adverbs appeared in the clause where the aspect word is detected, which are considered to modify the matched aspect word, therefore is an indication of the sentiment of the aspect mentioned.

In the review shown in Figure 2, there are three aspects detected: “storyline”, “acting” and “character”. For aspect “storyline”, the words that match the aspect wordset are “story” and “storyline”.

<table>
<thead>
<tr>
<th>Table 2: Detected Aspects and Their Wordset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspects</td>
</tr>
<tr>
<td>acting</td>
</tr>
<tr>
<td>directing</td>
</tr>
<tr>
<td>scenery</td>
</tr>
<tr>
<td>character</td>
</tr>
<tr>
<td>storyline</td>
</tr>
</tbody>
</table>


### Table 1: Core Aspect words

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>scenery</td>
<td>1.142</td>
</tr>
<tr>
<td>character</td>
<td>0.914</td>
</tr>
<tr>
<td>acting</td>
<td>0.904</td>
</tr>
<tr>
<td>storyline</td>
<td>0.830</td>
</tr>
<tr>
<td>directing</td>
<td>0.820</td>
</tr>
<tr>
<td>man</td>
<td>0.671</td>
</tr>
<tr>
<td>way</td>
<td>0.652</td>
</tr>
<tr>
<td>life</td>
<td>0.624</td>
</tr>
<tr>
<td>plot</td>
<td>0.621</td>
</tr>
<tr>
<td>comedy</td>
<td>0.509</td>
</tr>
</tbody>
</table>
I have several DVD movies based on “A Christmas Carol” story authored by Charles Dickens. This version is by far one of the best because it portrays the living circumstances in 19th Century England and has a clear storyline. Patrick Stewart, in his unique style of superb acting on the stage when he brilliantly performs in Wm. Character “Scrooge” is well portrayed in the movie. If the viewer is in search of a version that interfaces the most close to the message of writings authored by Charles Dickens, this is the best yet.

Figure 2: Aspect and Sentiment Words Detection

For “story”, there is no adjectives or adverbs detected in the clause, therefore is ignored. For “storyline” there are positive adjectives “best” and “clear” found in the clause. Therefore, for aspect “storyline” there are sentiment words “best” and “clear” detected. Similarly, for aspect “acting”, there are “superb” and “brilliantly” detected. For aspect “character”, there is no adjective nor adverb detected in the clause, therefore is ignored. To calculate the sentiment score of the adjectives and adverbs, the system uses SentiWordnet [2], which is an expansion of Wordnet [25] with sentiment score for each word. For example, word “unable” has sentiment score: positive: 0.0, negative: 0.75, objective: 0.25. Only positive and negative score is calculated in the consideration, however, adjectives or adverbs with objective value of 1 are ignored. For each aspect, the sentiment of which is calculated as average sentiment score of all the adjectives and adverbs detected to modify it. In the end, the score is normalized to a score in the range 1-5 (consistent with the scale of review on Amazon.com), and every review is represented by an aspect vector of (acting, directing, scenery, character, storyline). For example, this review has an aspect vector of (4.375, None, None, 3.23), “None” denotes that this aspect is not detected in the review. As to negation words, such as “not” and “don’t”, the system will negate the score and then normalized to the range 1-5 again. It is worth to note that not all reviews containing negation words should be negated, such as “not only” “but also”.

3.4 Review Clustering

As mentioned before, there is hardly a research in the opinion mining field which also performs clustering on the reviews to group reviews which share similar opinions together. However, different groups of people tend to have different opinions on the same product. For example, in the movie “Twilight”, some people think “it is so romantic and amazing”, however, the others think “sparkling vampire is a stupid thing”. Therefore it is necessary to cluster reviews which share similar opinions together, which may help the users find the reviews group where they fit in, and understand the product better than just read the top ranked reviews. Also, it is worth to mention that different products may have different numbers of clusters, some products are controversial, and therefore have more clusters, while some are not. To reduce the dimensionality of dataset and focus more on people’s opinions towards different aspect of the product, every review is represented as an aspect vector generated from the last phase, and the clustering is based on these aspect vectors.

There are several popular clustering methods which can be implemented in the prototype system. One popular method is partition-based clustering such as k-means [14]. However, k-means highly relies on the shape of clusters, clusters with a round shape perform the best, and it is also highly relying on euclidean distance as distance measurement. However, this prototype system can not use euclidean distance and many other distance measurements because there will be a lot of missing values in the dataset (not every aspect is mentioned by every review). It is also likely that the review clusters will not be a perfect shape of round and will have a lot of noise. Grid and hierarchical clustering are easily affected by noise, and have high time cost [33]. Therefore, in the end a density based approach such as DBSCAN [9] seems to be the best option. DBSCAN does not have requirement of the shape of dataset clusters, and is good at handling noise. In DBSCAN, two parameters need to be defined manually: eps and Minpts, eps refers to the distance to be searched around every point, Minpts refers to the minimum points in the distance eps of a point to form a core point. core points which are closed to each other will be connected together into one cluster. However, DBSCAN is very sensitive to the two parameters and is less good at handling less dense area. Therefore, an improved version of DBSCAN is chosen, which is called HDBSCAN [30].

In the clustering, each review is represented as an aspect vector generated from last phase. However, not every review will mention every aspect, therefore, almost every review has missing values in different aspects dimension. The most popular distance measurements such as euclidean distance can not be used. To handle this problem, this thesis designs a new distance measurement:

$$\text{Dis}(\text{review}_a, \text{review}_b) = \frac{\text{avg}(|\text{review}_a(i) - \text{review}_b(i)|)}{(\text{review}_a(i), \text{review}_b(i) \neq \text{None}, i \in (1, 5))}$$ (2)

In brief, the distance between two reviews is the average absolute value difference between each aspect scores of the reviews, “None” value is ignored.

To select the parameter eps, the prototype system tries to find the average number of nearest neighbor points of every points while increasing the distance value eps, when the number of neighbor points increases dramatically (the elbow point), the eps at this point is chosen [29] as shown in Figure3. To select the parameter Minpts, the system iterates through different Minpts values, and select the Minpts which has the highest silhouette score [31], which is a
measurement for clustering algorithms. In brief, if the points in a cluster are similar to each other, and not very similar to points in other clusters, then the silhouette score is high. In the end, every movie product has its own number of clusters of reviews to be presented to users.

4 EVALUATION

The evaluation process is divided into two parts. The first part is aspect extraction evaluation, which evaluates the quality of aspects generated by the prototype system in aspect detection phase and answers the first research question: is it possible to automatically extract aspects of products from reviews which are also consistent with manual modeling? In the aspect detection evaluation, by asking people who are knowledgeable in data modeling to model the movie products with important aspects, it is possible to check if the result is consistent with the system result. The second part is to evaluate the overall performance of the system, in which participants are asked to use both the reviews generated by the system and the reviews online to complete tasks. The purpose of this is to determine if the system indeed can help users understand products faster, and thus answer the second research question.

4.1 Aspect Extraction Evaluation

In the aspect extraction evaluation, it is divided into two parts. First, the performance of the prototype system in aspect detection is evaluated, i.e., how many aspects mentioned in the movie reviews are successfully detected? Second, the aspects generated by the system are compared with manual modeling, do people who have data modeling knowledge agree with the aspects generated by the system to be used to model the products? In brief, the evaluations first try to find out whether the system can successfully detect aspects of movies from reviews, then check if the aspects generated by the system are consistent with the manual modeling.

4.1.1 Aspects in Reviews. It is necessary to examine which aspects indeed appear in the reviews, and how many of which are successfully detected by the prototype system. The author manually read through reviews of movies under 5 genres (5 movies per genre, 5 top ranked helpful reviews per movie, 125 reviews in total). Then the author manually selects out which aspects of movies are mentioned in the reviews after reading, which are "storyline" "acting" "scenery" "character" "directing" "soundtrack" "humor". There are also some minor aspects mentioned by the reviews, such as "costume" and "location". However, they are not mentioned by the most reviews, therefore are ignored. Only the main aspects are listed here.

To compare the result with the system result, the system runs the aspect extraction on the 125 reviews. The prototype system manages to find the first 5 out of 7 of the aspects, i.e., "storyline" "acting" "scenery" "character" "directing". However, the system fails to find some other aspects which might be important, such as "soundtrack" "humor". The reason that this happens is because in the aspect extraction phase, when calculating the importance score of each word, the words "soundtrack" and "humor" have low similarity with the word "movie" in WordNet, even it has high TF-IDF score, in the end their calculated importance scores are low in the aspect extraction phase. However, in general, the performance of the prototype system is good, which manages to find most of the aspects mentioned in the reviews.

4.1.2 Aspects with Manual Modeling. The main purpose of this evaluation is to answer the first research question: do people with data modeling knowledge indeed agree with the aspects generated by the system to be used to model the movie products? i.e., are the system generated aspects consistent with manual modeling? However, other works do not actually carry out such a data modeling evaluation to ask people's opinions on product modeling, simply judging it themselves or use movie awards as aspects. For example, some works use the different awards of Oscar as aspect indicator of movies [34] (such as "cinematography" "music" "costume design"), or manually selects the aspects by several people without carrying a study including large amount of participants. The problem is, Oscar awards may be too professional and are not agreed by ordinary people. Moreover, some movies have low budget and do not have very fancy costume design but have excellent storylines, they are still highly ranked online. Also, selecting the aspects by several people may be biased. Therefore, it is necessary to carry out a study which actually asks a lot of people's opinions on which aspects should be chosen to model the movies. As mentioned before, the system uses a method which combines TF-IDF and similarity calculation to find aspect keywords of a product. The top 5 aspects found by the system are "acting" "character" "storyline" "directing" "scenery". These aspects will be compared with manual modeling result.

In the experiment, 180 students who are knowledgeable in data modeling are invited to complete a task. They first need to form a group of 2, which results in 90 groups in total. Then, each group will complete a task together, so that they can form some consensus over the task, partly eliminating individual bias (e.g. some individuals care about female appearance in the movie, while others not). In the experiment, each group is asked to model movie products by listing all the aspects of movies which they think are important in describing movies. They are also asked to give each aspect an...
importance score, (ranging from 1 to 5) as well as short explanations of why they choose it. An example of the respond from one group can be found in Table 3. As shown, the group thinks “acting” is the most important aspect, and gives an importance score 5. While “storyline” “scenes” have a bit lower importance score. “Explosions” has only a score of 2, it is because maybe only few people care about it. A distribution of the number of aspects that the participants use to model movie can be found in 4. As shown, most groups use 6 to 7 aspects to model the movie product.

In the analysis phase, the author notices that sometimes the participants do not use the same words to describe the same aspects, words such as “story” “plot” should be merged together under “storyline” aspect. Therefore, the author manually analyzes the explanations written by the participants, if the participants mean to write down the same aspect of a movie but using different words, then they will be merged under the same aspect. The total score of each aspect is the sum of all the importance scores of it, an average importance score of each aspect is also calculated, as well as the frequency of the aspects mentioned. In the end, the importance of each aspect is ranked by its total score. The total score, frequency of the top and bottom aspects mentioned, their average importance score from the experiment, and whether they are main aspects appearing in the reviews or are detected by the system are shown in Table 4.

In the table, the top 10 and the bottom 5 aspects from the experiment are shown. The bottom 5 aspects are not agreed by most of the groups, some of them contain very strange aspects such as “babes”, which refers to the female appearance in the movie. As mentioned before, some works use manual selection to find out the aspects of the products by several people, that’s why the works may be biased, because it may not be agreed by other people. Therefore, it is important to ask many people’s opinions on manual modeling of the product.

It is shown that the top 5 aspects mentioned by the participants also match with the results found by the system, i.e., “acting” “character” “storyline” “directing” “scenery”. The result indicates that the top 5 aspects found by the system are considered indeed descriptive in describing movie by the participants, and are also consistent with the manual modeling. Although the thesis focuses on movie products, it should be possible to be generalized to other products as well. Therefore, to answer the first research question, the evaluation indicates that it is possible to automatically extract aspects of online products from reviews which are also consistent with the manual modeling.

4.1.3 Generalization. To have a test whether the system can be generalized to products under other categories, the system is also tested with the reviews from “Cell Phone” and “Book” category. The top 5 aspects chosen for each of them are shown in Table 5. This work does not test the system in other domains thoroughly, however, from this rather simple result it still shows that the system has a potential in generalization.

4.2 System Evaluation

To evaluate the prototype system and answer the second research question: is it possible to develop an aspect-based extraction system which indeed can help users understand products faster? A user case study is conducted, where participants are invited to use the reviews generated by the system and the top 10 ranked helpful reviews of the corresponding movie product online, then complete tasks to answer questions related to the movie product. The purpose of it is to find out if the system indeed can help users understand products faster. Therefore, their performance in the task and the answers to the questions are analyzed.

4.2.1 Evaluation Setting. In the evaluation, 20 students from TU Delft are invited as participants. The reason of choosing 20 students is because the research [4] indicates that young educated people are the most loyal customers of shopping websites, therefore this evaluation may give a better insight. Each participant receives two types of reviews (top 10 ranked helpful reviews online and system generated reviews) of two different movies in printed version, which will help him understand two movies that he hasn’t watched before, then he fills in a questionnaire online (Appendix A). The evaluation is done by interviewing the participants one by one. To prove that the prototype system indeed can help people understand the product faster with less effort, the performance

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Importance Score</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Acting</td>
<td>5</td>
<td>The quality of the acting in a movie can make a great difference in the overall quality. Believable acting can be a great help to a movies perceived quality.</td>
</tr>
<tr>
<td>Originality of Storyline</td>
<td>3</td>
<td>Most big movies seem to have the same storyline with some characters changing names, they are predictable and bore a lot of people. A lot of people are therefore interested in movies with original storylines.</td>
</tr>
<tr>
<td>Amount of Explosions</td>
<td>2</td>
<td>A large amount of explosions will put of some viewers while others will really appreciate them.</td>
</tr>
<tr>
<td>Quality of Scenes</td>
<td>3</td>
<td>Simply put fight scenes can differ greatly in quality from the lameness that is Darth-Vader vs Palpatine to the awesomeness that is gypsy danger rocket punching a giant monster in the face. People that watch action movies want to see people get punched in an awesome way, not thrown over a railing for example. This can really make or break these kinds of movies for certain viewers.</td>
</tr>
</tbody>
</table>

Table 3: An Example of Aspect Task Respond
Table 4: Top and Bottom Results from Aspect Extraction Experiment

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Total Score</th>
<th>Average Importance</th>
<th>Frequency</th>
<th>Manual</th>
<th>System (125 reviews)</th>
<th>System (all reviews)</th>
</tr>
</thead>
<tbody>
<tr>
<td>storyline</td>
<td>346</td>
<td>4.67</td>
<td>74</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>acting</td>
<td>341</td>
<td>4.73</td>
<td>72</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>scenery</td>
<td>197</td>
<td>4.10</td>
<td>48</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>character</td>
<td>186</td>
<td>4.53</td>
<td>41</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>directing</td>
<td>177</td>
<td>4.43</td>
<td>40</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>sound</td>
<td>162</td>
<td>2.74</td>
<td>59</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>humor</td>
<td>160</td>
<td>3.01</td>
<td>53</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>originality</td>
<td>67</td>
<td>2.48</td>
<td>27</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>pace</td>
<td>49</td>
<td>3.06</td>
<td>16</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>cinematography</td>
<td>46</td>
<td>2.70</td>
<td>17</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>child friendliness</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>babes</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>morality</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>animal friendliness</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Disney level</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 5: Aspects from other categories

<table>
<thead>
<tr>
<th>Cell Phone</th>
<th>Books</th>
</tr>
</thead>
<tbody>
<tr>
<td>screen</td>
<td>story</td>
</tr>
<tr>
<td>battery</td>
<td>character</td>
</tr>
<tr>
<td>headset</td>
<td>novelty</td>
</tr>
<tr>
<td>charge</td>
<td>author</td>
</tr>
<tr>
<td>protector</td>
<td>end</td>
</tr>
</tbody>
</table>

Table 6: User Case Combinations

<table>
<thead>
<tr>
<th>Combination</th>
<th>Number</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR1 &amp; MS2</td>
<td>5</td>
<td>Ranked reviews for movie 1 first, then generated reviews for movie 2</td>
</tr>
<tr>
<td>MR2 &amp; MS1</td>
<td>5</td>
<td>Ranked reviews for movie 2 first, then generated reviews for movie 1</td>
</tr>
<tr>
<td>MS1 &amp; MR2</td>
<td>5</td>
<td>System reviews for movie 1 first, then ranked reviews for movie 2</td>
</tr>
<tr>
<td>MS2 &amp; MR1</td>
<td>5</td>
<td>System reviews for movie 2 first, then ranked reviews for movie 1</td>
</tr>
</tbody>
</table>

As mentioned, the evaluation chooses two movies which will be used in the evaluation tasks. The evaluation chooses movies with at least more than 100 reviews. The reason to choose this kind of movies is because the movies with very few viewers may already indicate that it is a bad movie, and therefore include not many controversial opinions in the reviews. To show the difference between top ranked reviews and system generated reviews, it is necessary to find movies with more controversial opinions. To find controversial movies, the controversial score of different movies is calculated, which is the ratio of number of positive reviews to number of negative reviews (movies have controversial score close to one are considered very controversial). However, it is very difficult to find movies with controversial score which is exactly 1 and have more than 100 reviews, the scores are normally higher than 4. In the end, Johnny Mnemonic (movie 1) and Seven Years in Tibet (movie 2) are chosen, which have controversial scores 4.2 and 4.6 respectively.

After reading each review, the participants will be asked to fill in a questionnaire which includes questions related to the movie products, their purchase decisions and their experience. The full questionnaire can be found in Appendix A. The reviews for participants to read can be found in Appendix B - E. When a participant

and answers of participants when using the system are compared with that of using top 10 ranked helpful reviews extracted from Amazon.com under the corresponding product. In the user case, each participant is required to read two types of reviews (ranked reviews and system reviews) for two movie product (denoted as M1 and M2) respectively. The ranked reviews are top 10 ranked helpful reviews from shopping website (denoted as R), the system reviews are review results generated by the system (denoted as S). For example, MR1 and MS2 refers to ranked reviews and system reviews of movie 1 respectively, with regard to Appendix B and C. Similarly, MR2 and MS1 refers to ranked reviews and system reviews of movie 2, with regard to Appendix D and E. The reason of using top 10 ranked helpful reviews is that it can simulate the real world situation, where people tend to consider the top ranked reviews before making purchase decision. The user case interview procedure is as follows, questionnaire can be found in Appendix A:

1. Explain procedures to participant
2. Participant answers demographic questions (Q1 - Q7)
3. Participant reads reviews of the first movie (timed)
4. Participant answers questions related to the first movie without reviews (Q8 - Q12, timed)
5. Participants answers questions related to their experience of the first task (Q13 - Q19)
6. Participant reads reviews of the second movie (timed)
7. Participant answers questions related to the second movie without reviews (Q8 - Q12, timed)
8. Participants answers questions related to their experience of the second task (Q13 - Q19)

As mentioned, the evaluation chooses two movies which will be used in the evaluation tasks. The evaluation chooses movies with at least more than 100 reviews. The reason to choose this kind of movies is because the movies with very few viewers may already indicate that it is a bad movie, and therefore include not many controversial opinions in the reviews. To show the difference between top ranked reviews and system generated reviews, it is necessary to find movies with more controversial opinions. To find controversial movies, the controversial score of different movies is calculated, which is the ratio of number of positive reviews to number of negative reviews (movies have controversial score close to one are considered very controversial). However, it is very difficult to find movies with controversial score which is exactly 1 and have more than 100 reviews, the score is normally higher than 4. In the end, Johnny Mnemonic (movie 1) and Seven Years in Tibet (movie 2) are chosen, which have controversial scores 4.2 and 4.6 respectively.

After reading each review, the participants will be asked to fill in a questionnaire which includes questions related to the movie products, their purchase decisions and their experience. The full questionnaire can be found in Appendix A. The reviews for participants to read can be found in Appendix B - E. When a participant
starts to fill in the questions which require him to list information related to the movie, the amount of time he spends will be timed. By asking the participant to fill in the information related to the movie (refers to questions 10-12 in the questionnaire) after reading the reviews, it is possible to monitor the memorability of the prototype system and their understanding of the products. It is assumed that if the system can help the users understand the products better, then they will be able to list more details related to the movie product, thus they remember more details related to the movie. However, the correctness of the understandings will not be examined, the reason is people tend to have their own preference, violent movies may be attractive to some people but not to others. Then the participants will be asked questions related to the satisfaction of their performance, difficulty of the tasks etc. The indexes chosen here are developed by NASA-TLX [12].

The reason of choosing two movies is to reduce the differences caused by a single movie (e.g. it may be easy for user to answer questions because the movie itself is controversial). The reason that the study asks every participant to read two types of reviews rather than dividing participants to two groups (each group reads only one type of review), is because people may have different understandings of the user case study and therefore spend different time in reading and writing. Putting the same participant in two different conditions can eliminate the time difference caused by individuals. However, the order of reading ranked ranked review (MR) first or system generated review (MS) first may also affect the outcome of the results (users may answer the questions of second movie faster because they get familiar with the settings of the experiment), therefore the order of the review reading is exchanged sometimes, which leaves 4 possible combinations. In the study, each combination includes 5 participants, as shown in Table 6. The total 20 participants recruited are master or phd students between 20-35 years old. According to the study of [4], educated young people are the most loyal customers of online shopping websites. Therefore, focusing on the students may give an insight of the performance of the most loyal customers of online shopping websites. Therefore, it is expected the answers of questions 10-18 will have a significant difference between “MR” and “MS”. The expected outcome is as follows:

1. Participants in “MS” feel less stress and more confidence in their purchase decision, less pressure and difficulty, and are more satisfied at their performance, in comparing with the ranked ranked reviews (MR). That is, it is expected the answers of questions 10-18 will have a significant difference between “MR” and “MS”. The expected outcome is as follows:

- Participants in “MS” spend less time in answering the questions related to movie after reading reviews.
- Participants in “MS” list more details related to the product (memorability)
- Participants in “MS” feel less mental stress and more confidence etc. (question 10-18)

4.2.2 Expected Outcome. It is expected that participants can answer questions related to the movie faster when using the system generated reviews (MS), listing more details related to the movie (memorability), feeling more confident in their purchase decision,
question 15. "average confidence" is how confident the participants feel about their purchase decision they make for question 16, this refers to question 17. "average helpfulness" refers to how helpful they think the review is in helping them, with regard to question 18. The detailed distributions of questions 13-18 can be found in the Appendix F.

As shown, there is no significant difference of the mental situation index (such as pressure, confidence etc.) between ranked review condition and system review condition (questions 10-18), the p-value is not significant in either 0.1 or 0.05 level. This means the participants do not feel significant difference with the help of ranked reviews and system reviews when completing the tasks, which is different from expected. There is no significant difference between the reading time of "MR" and "MS" neither, and participants spend even slightly longer time in reading in "MS". By observation, the reason that participants spend longer time in reading is because they do not understand the layout of system generated review (Appendix B - E). Therefore, they spend longer time in understanding the layout rather than actually reading.

However, there is a significant difference between the time the participants spend on completing the tasks of "MR" and "MS" (refers to questions 10-12). Averagely, the system improves the time of answering the questions for 44 seconds, which indicates that the system generated reviews do help participants understand the products faster. Also, the participants seem to be able to list more points related to the movie with the help of system reviews, averagely 1 more, with a significance p-value 0.0196, which indicates the system can help users understand and remember more details about the products. The author notices that sometimes participants spend more time in completing the tasks, because they have more points to write related to the movie. Therefore, it will be more helpful to take a look at the average time participants spend on listing every single point related to the movie. As we can see in the table, participants with the help of system reviews spend 11 seconds less on thinking and writing down per point, with a significant p-value 0.0096 in compare with the mean value of ranked reviews.

The user case study shows that the prototype system has a big potential in helping people understanding the products faster, remembering more details about the products. Even though the participants do not feel the difference of the reviews themselves when completing the tasks (judging from question 13-18), time-wise, they do perform much better with the help of system reviews than using ranked reviews. Even though the participants spend slightly longer when reading the reviews generated by the system, this is caused by the layout, it is believed that when the participants get used to the system generated reviews, they can read faster. There are some drawbacks of the experiment, the user case is not conducted on a real system, but with printed reviews generated from the system. Also, the sample is rather small, with only 20 participants and the system is focused on movie products. However, the study does show that the system has a big potential in helping people forming opinions fast on products. To answer the second research question, the evaluation indicates that it is possible to build a system to aggregate people's opinions on online products from reviews, which indeed helps users understand the products faster.

5 CONCLUSIONS

With the development of Internet, more and more people do shopping online. Shopping websites such as Amazon.com allow customers to leave reviews to the products after purchase, which help other people understand the products better. However, there are often hundreds or thousands of reviews under popular products, and the overview average scores of the products from the reviews provide not enough information to the users, which do not give information or ratings of different aspects of the products. If they simply read the top ranked helpful reviews, they will get biased opinions, because their preference may not match with the top voted or average opinions. To understand the products more without bias, the users need to read through a lot of reviews, which is tedious to do. There are works related to review summarization in opinion mining field, but some works do not consider different aspects of the products, simply classifying the reviews to negative/positive, some works require a lot of manual labeling as training data (not automatic), or ignore the conflicted opinions in the reviews by averaging all the opinions in the reviews. Under this circumstance, this thesis proposes two research questions: is it possible to automatically extract aspects of online products which are also consistent with the manual modeling of products in this domain? Is it possible to build a system to aggregate opinions on online products from reviews, which indeed can help users understand the products faster?

To answer these two questions, this thesis introduces a prototype system which is developed to automatically aggregate people's opinions from reviews on different aspects of online product, in order to help users understand the products faster. The prototype system is an aspect-based sentiment system. Then two evaluations are conducted: aspect extraction evaluation and system overall evaluation. In aspect extraction evaluation, the performance of the aspect extraction of the system is first evaluated. The aspects which appear in the reviews are manually selected out, and are compared with the aspects found by the system. Then a experiment is carried out, where participants who have knowledge in data modeling are asked to model the movie products with aspects they think are important, in order to determine if the system generated aspects are consistent with the manual modeling. In the system overall evaluation, participants are asked to use both reviews generated by the system and top 10 ranked reviews of the corresponding products to complete tasks, in order to determine if such system indeed can help users understand products faster. The dataset chosen in the prototype system is reviews and product description data extracted from Amazon.com, under Movie and Tv product category. However, the system should be also able to be generalized to products under other categories or on other shopping websites.

The prototype system is divided into three parts, aspect extraction, sentiment analysis, and review clustering. In aspect extraction, the system detects which aspect keywords should be used to describe the products based on the reviews of them, and find the words related to these aspect keywords to form aspect wordsets. The approach used is an importance score calculation method, which calculates a score for each word appears in the movie reviews, and the score is a combination of TF-IDF score and similarity score of the word with concept "movie" (based on WordNet path length
calculation). The TF-IDF score helps to find the most representative words in movie review domain, the similarity score helps to exclude the words not related to the movie themselves (such as delivery service, package complain). The top 5 ranked aspect keywords (acting, directing, characters, scenery, storyline) are selected as core aspect words. To evaluate the performance of aspect extraction, the main aspects that appear in the reviews are manually selected out, it turns out that the prototype system successfully find 5 out of 7 of them. Also, the aspects found by the system are consistent with the experiment, where 180 students who have knowledge of data modeling are asked to model the movie with the aspects that they think are important. The aspects found out by the system match with the experiment results. However, the aspects such as “humor” and “soundtrack” are failed to be detected by the system. The reason is because even these two aspect words have high TF-IDF score, they are not related to the concept “movie”, therefore they have low similarity scores and in the end have low importance scores. To find words related to these aspects keywords to form aspect wordset, the system still uses WordNet to find words that have high similarity with aspect keyword to generate aspect wordset for each aspect. If a word in a review matches with any word in the aspect wordset, then it is considered that this aspect is mentioned in the review. After that, the prototype system performs sentiment analysis in the clause where the aspect is detected, the adjectives and adverbs are used to calculate an average sentiment score for the aspect based on SentiWordNet (score out of 5), which is a dictionary with sentiment value for each word. After sentiment analysis phase, each review is calculated with a vector of 5 aspect score. To cluster the reviews which share similar opinions on the product, the system performs HDBSCAN based on the vectors of reviews, and present the resulted clusters to users.

In the end, the thesis introduces two evaluations: aspect extraction evaluation and system evaluation. In the aspect evaluation, 180 participants with topic modeling knowledge are invited to model the movie products with aspects they think are important (form a group of 2, in total 90 groups). Also, the reviews are manually analyzed examined to find out which main aspects actually appear in the reviews. The result shows that the aspects detected by the prototype system are consistent with the manual modeling. The system manages to detect 5 out of 7 aspects which actually appeared in the reviews, and these 5 aspects are also the top 5 aspects considered important by the 180 participants. To have a test on the generalization of the system, the system is also tested with the reviews from products of other two categories “Cell Phone” and “Books”. The results show that even though the system is not perfect, it still manages to find some important aspects of these two aspects, therefore has a big potential in generalization. In the system evaluation, a user case study is conducted to evaluate the prototype system, where each participant is asked to complete two tasks one by one: read two different reviews (ranked reviews and system reviews) of two different movies. Then the participant is asked to answer questions related to the movies, such as advantages/disadvantages/controversial parts of the movie, or their mental situation, purchase confidence etc. The time participants spend in the experiment is timed (reading, answering questions are timed separately). The result shows that even though the participants do not feel the difference of their performance themselves in two tasks, the system reviews do help them complete the tasks faster, and they are able to remember more details related to the movie products with the help of system reviews. There are several drawbacks of the evaluation setting, such as the user study sample is rather small, and the study is not conducted on a real system, but with printed out reviews generated by the system. However, it still shows a big potential in the prototype system, that it can help users understand the products better, remembering more details.

Finally, to answer the two research questions: the first evaluation shows that it is possible to automatically extract aspects of online products which are also consistent with the manual modeling of movie products in this domain, the second evaluation shows that it is possible to develop a system which aggregates opinions on online products from reviews and helps users understand faster. Even though the system and evaluation are done in the field of movie products on Amazon.com, they should be able to be generalized to products under other categories or on other websites as well. The thesis shows there is big potential in this kind of system.
APPENDIX
A QUESTIONNAIRE

Scenario: you are deciding whether you want to buy two DVD movie products, you will be given review information of this 2 movies (each movie’s review is represented in different form). Based on these reviews, you will be asked to answer several questions related to the movies.

Before Task
1. How good is your English?
   not fluent 1 2 3 4 5 6 7 very fluent

2. What is your age?

3. What is your education level?
   1. High school or lower
   2. Bachelor
   3. Master
   4. PhD or higher

4. What is your gender?
   1. Male
   2. Female
   3. Prefer not to specify

5. What is your nationality?

6. How often do you watch movies?
   1. Once or more in a day
   2. Once or more in a week
   3. Once or more in a month
   4. Once or more in a year
   5. Less than once a year

7. What types of movie do you like? (one or more)
   1. Action
   2. Animation
   3. Comedy
   4. Documentary
   5. Adventure
   6. Biography
   7. Crime
   8. Romance
   9. Fantasy
   10. Horror
   11. Science fiction

Task: Movie 1 or Movie 2 (depends on the order)

Your task is to decide whether you want to buy the DVD product of this movie, you are provided with the reviews from the former customers, try to find the strong/weak/controversial points of the movie when you read the reviews. After reading the reviews, you will not be able to read the reviews again when answering the questions.

8. Have you heard of this movie?
   1. Yes
   2. No

9. Have you watched this movie?
   1. Yes
   2. No

10. List 2-3 strong points of the movie (bullet points are OK)
    e.g. acting is awesome, music is nice

11. List 2-3 weak points of the movie (bullet points are OK)
    e.g. story is bad, acting is poor

12. List 2-3 points of the movie that reviewers have different opinions about (bullet points are OK)
    e.g. some reviews think acting is good, some reviews think it is bad

After Task
13. How much time pressure did you feel in completing the task? (in summarizing points related the movie)
   not stressful 1 2 3 4 5 6 7 very stressful

14. How satisfied are you with your performance in the task? (in summarizing points related the movie)
   not satisfied 1 2 3 4 5 6 7 very satisfied

15. How hard is this task? (in summarizing points related the movie)
   not hard 1 2 3 4 5 6 7 very hard

16. How likely is it that you would purchase this movie DVD after reading the reviews? (ignoring financial situation)
   definitely would not 1 2 3 4 5 6 7 definitely would

17. How confident are you in the purchase decision you made just now?
   not confident 1 2 3 4 5 6 7 very confident

18. How helpful are the given reviews in helping you form opinions about the movie?
   not helpful 1 2 3 4 5 6 7 very helpful

19. Do you want to leave any comments related to this task? (skip if no comments)
B MR1: RANKED REVIEW OF MOVIE 1: JOHNNY MnEMONIC

1. Yes, you have to suspend your disbelief in certain areas, the plot for example. Yes, doubling 80 gigs to 160 gigs is laughable with the mega gigs available today (the movie was made over 15 years ago). Yes, Keanu Reeves acting is wooden (when is it not?)

2. It’s a sci-fi. If you like sci-fi, you’ll like this. Yeah, parts of acting are cheesy... but again... it’s a “techie sci-fi”. ;)

3. Aside from a few technological items that were questionable, Johnny Mnemonic was not nearly as “dated” as I thought it might be. I used to enjoy this movie on laser disc, but hadn’t watched it since I put away my laser disc player a few years back. Given our current world, with Trump in charge in Washington, the Pharmaceutical Companies doing what the movie describes, and technology less than a decade away from routine enhancement implants, I found it very engaging, and VERY relevant. It’s even likely that lack of guaranteed security on the internet will make such Mnemonic couriers as “Johnny” useful if not necessary. I needed its hopeful ending which did not disappoint.

4. This movie is living proof of his genius. Words cannot express the violent emotions and physical trauma Reeves’s head must be going through when he pumps 80 GB of data into his brain. When Henry Rollins tells him that his brain is leaking data (or something like that), tensions run incredibly high in this cyberpunk masterpiece. Reeves delivers an oscar-worthy performance in Johnny Mnemonic putting him in the same arena as Hopkins, Pacino, and DeNiro.

5. I watched this for a college course, and while enjoyable, it’s not that amazing compared to the actual short story. I feel like the whole theme of sticking it to the one company holding a patent on the cure for the disease everyone has, was missed in the movie. The movie didn’t make me feel like they were trying to undermine to monopoly.

6. This movie is better than people tend to think. Keanu does a good job, as do Dina Meyer and Henry Rollins. Ice-T plays basically the same character he does in everything, but it fits pretty nicely in this movie. Overall, it’s a good adaptation of Gibson’s short story, with the seemingly required reworking in places (no Killing Floor, sadly). However, it remains a good storytelling of the cyberpunk genre, and it should essentially be required viewing for Shadowrun players, as the movie’s setting is the best visual treatment we have so far, even minus the magical component of Shadowrun. If you have the fullscreen DVD version (and I’ve never personally seen a widescreen version, despite looking for one for years) and you want the widescreen, this is your best bet.


8. I love this movie. I accidentally found it on my xbox 360 app and am glad that I now own it. It is the perfect 80s movie and I watch it over and over. It just sets me in a good mood every time.

9. When I heard that my husband had never seen Johnny Mnemonic, well that was enough for me. I immediately ordered him a surprise gift on blu-ray. What a great movie. We thoroughly enjoyed it. Ice-T - awesome. Jones - who doesn’t love dolphins? And we both wish we had laser fingernail lassos.

10. This could have been a much better film if they got a decent actor to play Johnny. Keanu just ain’t cuttin it.
C  MS1: SYSTEM REVIEW OF MOVIE 1: JOHNNY MNEMONIC

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Rating</th>
<th>Keywords</th>
<th>Aspects</th>
<th>Rating</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acting</td>
<td>3.5</td>
<td>Likeable, nice, raging, large, sudden</td>
<td>Acting</td>
<td>3.0</td>
<td>Wooden, bad, disappointing, embarrassed</td>
</tr>
<tr>
<td>Directing</td>
<td>None</td>
<td>None</td>
<td>Directing</td>
<td>2.1</td>
<td>terrible</td>
</tr>
<tr>
<td>Scene</td>
<td>3.9</td>
<td>Good, honorable, thick, original</td>
<td>Scene</td>
<td>2.7</td>
<td>Stupid</td>
</tr>
<tr>
<td>Character</td>
<td>3.7</td>
<td>charismatic</td>
<td>Character</td>
<td>3.0</td>
<td>Unimaginative, blunt, spoiled, fair</td>
</tr>
<tr>
<td>Storyline</td>
<td>3.8</td>
<td>Valuable, engaging, short, right, amazing</td>
<td>Storyline</td>
<td>3.3</td>
<td>Short, ridiculous, absurd, original</td>
</tr>
</tbody>
</table>

1. Great movie about a world in which technology becomes the enemy in the form of the shakes. The people with the cure don't want you to have it and there is a genetically enhanced Dolf Lundgren trying to cut your head-off. Your head has built in memory storage but you overloaded it with a very large save now you need help from some quack doctor to save not only yourself but also humanity.

2. This movie is pure cyberpunk, and carries all of the elements of what cyberpunk fans look for in that type of movie. You have mega-corporations and crime lords that control the world, cybernetically modified street punks and corporate mercenaries, three dimensional neural computer interfaces (which was quite interesting considering the Internet was just blossoming when the movie came out), outcast societies and street fixers, you name it. It all brings back to me nostalgia from reading books like Neuromancer and the Shadowrun series. Amazing story.

3. This probably my favorite Keanu Reeves flicks. The production quality isn't as good as the matrix series, but the story is more intriguing (IMHO).

4. People with hi-tech body implants, cyberpunks galore and some awful sickness induced by a technology gone wild, all situated in a world where computers, drugs, rust and dirt are the only winners. William Gibson's world is there, perfectly pictured in this superb movie. I recommend this movie.

5. This a pretty good action movie, futuristic in nature, with a storyline that is OK. I enjoyed it. Does have some bloody scenes if you are concerned about that sort of venue but overall it is a fun movie to watch. It is in the mode of "Blade Runner," "Logan's Run," and others of the genre. If you liked those you will like this movie.

1. The title sums it up, it's a good attempt at what could happen in the not too distant future, but it's not as imaginative and ground-breaking as The Matrix. The movie does give his patented another view of reality, but it's what could happen, not another view on what is happening. Like I said it's a good DVD movie, but I wouldn't want to pay $12.00 to see this.

2. The basic storyline is quite interesting but the story itself is less than interesting. The film has potential it just is never realized.

3. I saw this AWFUL excuse for a movie in the theater in the Spring of 1995, and to this day I consider it one of the WORST movies I have ever seen (to date) and an ABSOLUTE waste of the cost of the ticket.I don't even remember now (it being over a decade later) why I even went to see it. But I do remember how stupid and boring it was and how I kept dozing off throughout this sorry excuse for a film. Looking at some of the other reviews I can see that there are others who consider Johnny Mnemonic one of Keanu Reeves's biggest flops.I like Keanu Reeves, but he should be embarrassed for being part of such a waste of people's money and time! If I could give this movie NO STARS, or NEGATIVE STARS, I would. But I can't on this site, so I'm giving it 1 star, though it deserves much less! 1 STAR!!!!!!!

4. A disappointing treatment of a Cyberpunk classic, Johnny Mnemonic tries too hard and in doing so completely fails to impress. The dark, oppressive atmosphere of William Gibson's future comes across rather well only to be spoiled by blunt and unimaginative acting. The special effect sequences suffer the same fate and come across more like a poor advertisement for the land of tomorrow than an accurate depiction of Gibson's view of the Internet of the future. All in all the movie's a real let down and there's more to be gained by reading Gibson's Cyberpunk novels: Neuromancer, Count Zero, Mona Lisa Overdrive and Burning Chrome (the short story collection from which Johnny Mnemonic was taken).

5. This film has an interesting setting and background. However, the film fails to realize just how cool of a premise it could be and we are thrust into a laughable story with corny dialogue and bad acting. Dolph Lundgren makes a fool of himself as a mercenary priest out to kill main-character Johnny. The action in this movie fails to generate anything exciting and the script fails to give us characters to care for.
<table>
<thead>
<tr>
<th>MR2: RANKED REVIEW OF MOVIE 2: SEVEN YEARS IN TIBET</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.</strong> I truly enjoyed this movie as it covered many momentous events in history, from WWII to the Chinese occupation of Tibet. The scenes in Tibet itself were gorgeous, with wonderful costumes. It was interesting to watch Henrich change from an arrogant man to a humble one. His mentorship of and friendship with the young Dalai Lama was heartwarming. In the end, I was happy to see that Henrich’s own son accepted him and even took up mountain climbing too. What I didn’t find very believable is that Brad Pitt (and David Thewlis, for that matter) didn’t age a bit during those years in the prison camp AND the seven years in Tibet. The make-up artists should have made them age realistically. (Also living in a dry climate will also speed the aging process along.) At the end of the movie, Henrich didn’t seem old enough to have a teenage son!</td>
</tr>
<tr>
<td><strong>2.</strong> This film depicts spellbinding history and drama with a great cast, beautiful scenery and excellent costuming. It piqued my interest in Tibet in which case I was compelled to research and explore. It’s not often a film urges me to delve into its background. I highly recommend this film.</td>
</tr>
<tr>
<td><strong>3.</strong> Excellent movie!!! I enjoyed seeing Brad Pitt’s transformation from a self centered thrill seeker, to a compassionate, caring, thoughtful individual! Amazing!!!</td>
</tr>
<tr>
<td><strong>4.</strong> Way more than I expected. Based on a true story, it renewed my faith that being on the Journey is the most important thing. I had no idea about this real life back story of the Dali Lama as a young man, and his patient healing of an Austrian Man. What the movie meant to me can not be captured in words.</td>
</tr>
<tr>
<td><strong>5.</strong> Great Great Movie! I think everyone should see this! ... And not just because Brad Pitt is a babe! It moved me and I thought about it for days afterwards.</td>
</tr>
<tr>
<td><strong>6.</strong> A selfish man leaves his pregnant wife a few weeks before she is about to give birth to their first child, in order to climb a mountain for his country. During the climb he and the others are arrested by the British and are kept in a prisoner of war camp. After escaping and trekking for miles, he and a fellow prisoner end up becoming the first foreigners to enter Tibet. The story continues on as his relationship with the Dali Lama affects who he is deep down and he becomes a better man.</td>
</tr>
<tr>
<td><strong>7.</strong> How many times have you heard this: it is not as good as the book. As is usually the case with books made into movies, it omits or underplays some parts. The book gave me a glimpse into Tibet before the Chinese came in; the movie lost a lot of that. The movie also seemed more obviously political, though from what I have heard the Dalai Lama approves. If you like the movie, I strongly recommend you read the book.</td>
</tr>
<tr>
<td><strong>8.</strong> Really long drawn out story. The description of the movie is not accurate. It felt like seven years. The good part was sensing how the main character change. The change had little to do with the Dalai Lama. Life’s circumstances and insight about his own decisions/choices changed him.</td>
</tr>
<tr>
<td><strong>9.</strong> Fascinating true story engaging the mind about will, determination, and focus...all relating to two men in a different culture and how they learn to know and respect each other. Western culture meets Tibetan culture and lifestyle, based on faith in its spiritual leader. An interesting way to educate others not only about respect and tolerance for others, but most importantly, learning appreciation and admiration for differences.</td>
</tr>
<tr>
<td><strong>10.</strong> It was boring. There was not enough interaction with the Dalai Lama.</td>
</tr>
</tbody>
</table>
Summary 1

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Rating</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acting</td>
<td>3.8</td>
<td>Spiritual, actually, best, compassionate</td>
</tr>
<tr>
<td>Directing</td>
<td>3.2</td>
<td>Decent</td>
</tr>
<tr>
<td>Scene</td>
<td>4.1</td>
<td>Paternalistic, great excellent, interesting, superb</td>
</tr>
<tr>
<td>Character</td>
<td>3.7</td>
<td>Quiet, must, great, even, excellent</td>
</tr>
<tr>
<td>Storyline</td>
<td>3.8</td>
<td>True, stunning, difficult, engaging, exotic</td>
</tr>
</tbody>
</table>

1. Stunning visuals and an engaging story make this film a winner. Blue-ray makes it even better, as some scenes will take your breath away with grandeur and beauty.

2. Excellent film great acting and direction - shot on location made it very special and then it is based on a true story that made a powerful political statement on China’s dominant rule.

3. I had been avoiding this movie until I finally broke down, bought a DVD player and rented :Seven Years In Tibet:. Brad Pitt was the reason I had been avoiding it, and I’m here to say right now that was stupid of me. He may only be an adequate actor at best, but this is a fine tale, coupled with mesmerizing cinematography and a noble portrait of the Tibetan people and their spiritual leader, the Dali Lama.

4. Encouraged me to read the book and view the movie Kodon. Actually, I enjoyed Seven Years in Tibet so much, that I prompted my wife to see it, and appreciated it even more the second time. Really deserves a five star rating.

5. This is a great story about humankind; people learning to know themselves, serve each other, learn from each other, become more spiritual. It is beautifully filmed. Apparently it is correct and the main character remained friends with the Dali Lama until his death. Uplifting and spiritual.

---

Summary 2

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Rating</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acting</td>
<td>3.0</td>
<td>Bad, arrogant, fair, average</td>
</tr>
<tr>
<td>Directing</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Scene</td>
<td>3.5</td>
<td>Amazing, impressive, better</td>
</tr>
<tr>
<td>Character</td>
<td>3.4</td>
<td>Nice, interesting</td>
</tr>
<tr>
<td>Storyline</td>
<td>2.8</td>
<td>Boring, not interesting, slow, long</td>
</tr>
</tbody>
</table>

1. This is a great story that was poorly told in this movie. I felt like I was watching Brad Pitt in Tibet which is a testament to his poor acting skill. I really wanted to like this movie but was thoroughly disappointed. Even a soundtrack by Yo-Yo Ma could not save it. The directing was also distracting with poor pacing and choppy editing. I really wish this had been a great movie and I was rooting for it but it fell flat I am very sorry to say.

2. The story was interesting, but there are several parts that don’t make immediate sense. You have to just wait a moment and it gets explained soon after the confusing scene.

3. This movie was so boring. The scenery was okay, but the rest of the movie was horrible. And as for that fake Australian accent...what was Brad Pitt thinking?

4. Being interested in fully understanding the issues related to Tibet, I thought I’d give this a try. I see clearly that I’m going to have to go to other documentaries for information. Although there was redeeming information, this was a Hollywood-ized version of Heinrich Harrer’s experience. Brad Pitt was not believable in this role. Interesting story, but the movie was really long and slow. Glad I watched it once, but probably wouldn’t watch it again.

5. How many times have you heard this: it is not as good as the book. As is usually the case with books made into movies, it omits or underplays some parts. The book gave me a glimpse into Tibet before the Chinese came in; the movie lost a lot of that. The movie also seemed more obviously political, though from what I have heard the Dali Lama approves. If you like the movie, I strongly recommend you read the book. Really long drawn out story. The description of the movie is not accurate. It felt like seven years. The good part was sensing how the main character change. The change had little to do with the dali lama. Life’s circumstances and insight about his own decisions/choices changed him.
QUESTION DISTRIBUTION DIFFERENCES BETWEEN MR AND MS

No.13: Pressure of MR task

No.13: Pressure of MS task

No.14: Hardness of MR task

No.14: Hardness of MS task

No.15: Satisfaction of MR task

No.15: Satisfaction of MS task

No.17: Confidence of MR task

No.17: Confidence of MS task

No.18: Helpfulness of MR task

No.18: Helpfulness of MS task