Identifying tour structures in freight transport by mining of large trip databases

Nadi , Ali; van Lint, Hans; Tavasszy, Lorant; Snelder, Maaike

DOI
10.1109/ITSC45102.2020.9294432

Publication date
2020

Document Version
Final published version

Published in
2020 IEEE 23rd International Conference on Intelligent Transportation Systems, ITSC 2020

Identifying tour structures in freight transport by mining of large trip databases.*

Ali Nadi, Hans Van Lint, Lóránt Tavasszy, Maaike Snelder *Member , IEEE*

Abstract— **Scheduling and Routing in freight transport are usually the end products of an optimization process. However, the results may differ due to the heterogeneity of rules in different transport markets. Since the understanding of these decision rules is important for disaggregate freight modeling, this paper investigates the development of an effective decision tree method for extracting them from an extensive freight transport data. We applied the method to model departure time and type of tours in freight transport of agricultural products. Having these two models together help us understand the whole anatomy of the freight activities for the selected transport segment. The models highlight the characteristics of time-of-day freight activities for this sector and indicate the importance of spatial and temporal characteristics in capturing the distinctions of the type of tours.**

I. Introduction

Identifying the structure of tours in freight activities is crucial for freight and traffic management. Time-of-day (i.e. scheduling) and type-of-tours (i.e. Routing) are the two key characteristics of the freight transport activities that are believed to be conditioned upon the motorways' level of service [1]. From a travel demand management perspective, Time of day models shed light on the commercial vehicles pattern to understand their sensitivity to congestion.

The importance of time-of-day modeling is even more relevant in the light of the increase in containers' throughput of large logistic hubs such as port of Rotterdam (roughly 5 % according to [2]). Dynamics in departure time of freight traffic may have large impacts on motorway congestion; whereas in turn, those congestion problems may have large impacts on logistics operations [3]. Despite its importance, there is little research on freight transport departure time choice. Many of the earliest freight demand modeling frameworks use Monte Carlo simulation in which departure times are sampled from (limited) observed start times [4]. Probably, the first time-of-day model in freight transport is proposed in reference [5]. In this model, the authors use a discrete choice model estimated on survey data to obtain the probability that a delivery tour departs at a certain time of day from an origin. A recent time-period choice model based on stated preferences for road freight transport can be found in [6]. This research is followed by [7] using revealed preference (GPS tracking data of trailer) to model time of day choice. These authors use a nested logit model to improve the models' fit. A discrete-continuous Probit model is also proposed in [1] for modeling the time-of-day choice behavior of commercial vehicles in urban areas.

The type of tour is important for transport managers to understand how logistic operational strategies change due to the conditions on motorways. For example, they can identify whether reducing/increasing the number of trips per tour (i.e. intermediate stops) can improve the efficiency of commercial vehicle tours when facing congested areas. There is a limited amount of literature that focuses on "type of tours" modeling as well. An early study of commercial vehicle delivery strategy can be found in [8], which is followed and modified further by [9]. These studies used mixed and multinomial logit to model multiple types of tour classification. Researchers in [10] develop a multiple discrete-continuous choice method to model a joint distribution of type of tour and number of trips for commercial vehicles.

From the literature, it is clear that modeling time-of-day and type of tour in freight transport is a complex task, particularly since the freight transport industry includes many submarkets. Examples are agricultural products, food products, and chemical products. Generally speaking, they all have the same pick-up and delivery scheduling tasks to satisfy their demand. However, their performance is quite diverse when it comes to the operational details. Due to this heterogeneity, understanding and simulating the disaggregate freight transport activity is an important topic in freight modeling. One of the challenges is that the data for different industrial sectors are not easily available. Another key issue is that classical choice models can be used to study the preferences of decision-makers when he/she can make one choice among the discrete alternatives. However, this is not a good assumption to model type-of-tour and time-of-day departure time in freight transport. Because these features are the output of an optimization process (i.e. routing and scheduling) and thus rather more associative rules than the discrete choice of decision-makers.

In this research, we propose an alternative *machine learning* approach to model departure times. We learn these patterns from a unique and extensive database of tours, from which we derive decision rules to explain tour patterns in the freight transport activities. In this paper, we present two analyses. The first part considers descriptive scheduling of

* This research is supported by a grant from the Netherlands Organization for Scientific Research (NWO) for the ToGRIP project.

Ali Nadi is with the Delft University of Technology, Department of Transport and Planning, Stevinweg 1, 2628 CN, The Netherlands (corresponding author to provide phone: 031-626469-382; e-mail: a.nadinajafabadi@tudelft.nl).

Hans Van Lint is with the Delft University of Technology, Department of Transport and Planning, Stevinweg 1, 2628 CN, The Netherlands (e-mail: J.W.C.vanLint@tudelft.nl).

Lóránt Tavasszy is with the Delft University of Technology, Department of Transport and Planning, Stevinweg 1, 2628 CN, The Netherlands (e-mail: L.A.Tavasszy@tudelft.nl).

Maaike Snelder, is with the Delft University of Technology, Department of Transport and Planning, Stevinweg 1, 2628 CN, The Netherlands and TNO, Anna van Buerenplein 1, 2595 DA, The Hague, The Netherlands (e-mail: M.Snelder@tudelft.nl).

tours using decision trees which aims to find patterns in the departure time of tours. In the second part, we extract probabilistic rules that can explain different pick-up and delivery strategies and predict the average number of stops simultaneously. These two models together explain "the anatomy of tours" taking both routing and scheduling into account. The specific contributions of this paper are: (a) analytical insight of daily activities in freight transport from empirical data with many detailed features, (b) a decision tree classification approach to classify tours according to their scheduling and routing. (c) a method to fuse loop detector data and ground truth tour data to analyze the efficiency of tours.

The remainder of this paper is organized as follows: Section II describes the data structure and fusion of multiple data sources. Section III presents the methodology with which we characterize freight activities. Section IV. denotes results of the descriptive tour scheduling model and marks findings from descriptive analysis of the freight routing patterns. Section V, in the end, offers concluding insights.

II. Materials

For this study, we make use of multiple data sources to enrich one from another. The datasets we use are first the XML data set of truck diaries provided by the Central Bureau of statistics in the Netherland (CBS). In total CBS collected 2.65 million shipment records from the year 2015 which contain information regarding loading/unloading locations, commodity type, and vehicles used. The database has also been used in the Multi-agent simulation system for urban good transport (MASS-GT) models [11] The second data set is the distribution center (DC) database which comes from Rijkwaterstaat (RWS) in the Netherlands and contains over 1600 registered distribution centers along with their 6-digit postcode, size, and sectors. Third, we use the transshipment terminal (TT) dataset which is provided by IDVV-binnenvaart game (IDVV inland shipping game) and contains information regarding their postcode, size, and annual throughput. The fourth data set we use for this research is vehicle counts. In the Netherlands, the National Data Warehouse (NDW) provides a stream of vehicle counts collected from loop-detectors on motorways.

A. Data Fusion

The XML data contain the geographic location of trucks loading and unloading along with detailed information about other tours and trip characteristics. The Data however lack the intermediate stop characteristics. As this information is important for understanding the tour pattern, we used a hierarchical multi-step deterministic and probabilistic matching algorithm to characterize intermediate stops in XML data with DC and TT data. More information about the data fusion process is in [12].

B. Count Data Processing

Dispatchers in freight transport companies usually use optimization software to plan the tours. This software usually takes expected travel time into account while making routing and scheduling tables. This can affect the sensitivity of tours to congestion. To understand how the perception of congestion can reflect on characteristics of tours in different sectors, we make use of count data to obtain this perception of congestion for the intermediate stops. To estimate the aggregate congestion level for each geographical zone, we used the same method proposed in [13].

C. Congestion Indicator

We use a moving average of delays encountered as the congestion indicator. To calculate this delay, we first require average speed on a road link. We then calculate the average speed as a moving average of speeds over a period T for the link i.

$$\bar{v}_{i,T} = 1/T \sum_{j \in [t_0, t]} v_{i,j} \tag{1}$$

Where T is the duration of the period, t_0 and t are the first and last time slots in T respectively, and $v_{i,j}$ is the average speed obtained from loop detectors on the link i for the period T. while we considered the maximum measured average speed over a period as the free flow speed v^{free}, we define v^{min} as the lowest average speed for the link i:

$$v_{\min} = \min \bar{v}_{i,T} \tag{2}$$

Then the average delay during period T for the link i is:

$$d_{i,T} = 60 \Big/ (1/v^{\min} - 1/v^{free}) \tag{3}$$

We use $d_{i,T}$ as an indicator for the congestion level for period T on link i measured in minutes per kilometer. Modifying the T parameter can limit the estimation of delay for a specific period. We used three specifications for this parameter i.e. morning peak (7:00 -10:00 AM), evening peak hour (3:00 to 7:00 PM), and rest of the day. We calculate the aggregate congestion level (CL) for every zone (PC4 postcode) based on the average delay indicator:

$$CL_{z,T} = \sum_{i \in z} l_i d_{i,T} \Big/ \sum_{i \in z} l_i \tag{4}$$

The threshold for this indicator is considered to be 10 seconds per kilometer in peak period as is recommended in [13] for motorways in Netherland.

To make use of this indicator in our tour type and time of day analysis, first, we calculate the proximity of every zone to the congested zones, this way less congested zones that are close to or surrounded by high congested zones are also considered as congested. Secondly, we assumed that every vehicle takes the shortest path from origins to destinations because the actual path and time of arrival for the intermediate stops are not reported in XML data. The only reported timestamps in this data belongs to the start and end of the tour. This is however a fair assumption as our congestion indicator takes a wide range of periods so that the arrival time of a trip would fit to the correct periods with high probability regardless of the actual path. Finally, we define two binary variables as if the first and later pick-up or delivery locations are in a congested zone at the time of arrival. This gives us information on the variation of freight activity patterns based on the perception of the congestion.

TABLE I. LIST OF EXPLANATORY AND TARGET VARIABLES

No.	*Variable*	*Description*
E1	Number of shipments in the first visited location.	0: 0-2 1: 2-7 3: > 7
E2	Day of week	[0 6]: [Mon. Sun.]
E3	Visit to Distribution center	1: yes 0: no
E4	Visit to transshipment terminal	1: yes 0: not
E5	Congestion state of first intermediate stop	1: congested zone 0: otherwise
E6	Congestion state of other intermediate stops	1: congested zone 0: otherwise
E7	Average Load factor	0: 0-0.3 1:0.3-0.4 3: >0.4
E8	Tour distance	0: 0-47 km,1: >47
E9	Average Distance of trips	0: 0-37 km 1: 37-64 km 2: 64-95 km 3: 95-140 km
E10	Empty container/pallets	1: if a trip is with an empty container or pallets 0: otherwise
E11	Vehicle Type	0: truck 1: truck trailer
T1	Type of Tour	1: direct 2: collection 3: Distribution
T2	Number of stop	Min: 1 max:33
T3	Departure time	1 : (6:00-11:00] 2 : (11:00-15:00] 3 : (15:00-20:00] 4 : (20:00-6:00]

D. Tour data

Table 1. shows a description of the explanatory (labeled with E) and target (labeled with T) variables that we used for our analysis. The continuous explanatory variables E1, E7, and E8 are discretized into 4 categories based on their distribution quantiles.

III. METHODS

The decision tree is one of the simplest but powerful machine learning methods to build descriptive models. These models identify which covariates can explain the variability of the response variable by recursive partitioning of all the data according to the most significant covariate [14]. There are three strong reasons for the popularity of decision trees. First is the interpretability of the tree structures. Many machine learning algorithms have high prediction accuracy but are essentially black boxes. Decision tree models, however, have the joint advantage of being interpretable and having high accuracy. The second advantage of decision tree models is that these methods make no probability distribution assumptions but are still able to identify explanatory variables and detect interactions among them. There are different types of decision tree methods. The first regression tree algorithm is Automatic Interaction detection (AID). This algorithm was further improved by [15] through Chi-squared Automatic Interaction Detection (CHAID). However, the disadvantages of these methods include sensitivity to the overfitting problem as well as a bias towards covariates in case of many split possibilities [16]. To address this problem, an unbiased recursive partitioning by conditional partitioning (Ctree) was proposed in [16]. In this paper, we utilize Ctree to model both departure time and type of tours. Besides the general advantages of decision trees, this decision-tree method does not require any pruning and thus provides us with a more robust model with less complex rules. As in our type-of-tour model, we want to model a joint discrete type of tour and continuous distribution of the number of stop for each tour type, we add one preliminary step to the method to be able to handle the bivariate discrete-continuous response variable.

A. Discretizing continuous variable

To handle bivariate discrete-continuous variables, there are three regular approaches. One is to model each response variable separately which does not take into account the joint distribution of the two random variables. The second approach is to use copulas joint distribution to make a joint distribution from the bivariate variable. The third approach is to discretize the continuous variable and coupling the two discrete variables for all possible matches. For this research, we used the third approach as it is simple to implement and takes into account the correlation between the two response variables as well. We use however a systematic way of discretizing the continuous response variable. For each discrete value in the discrete response variable, we take the probability distribution of the continuous variable. Then for each category, we use the k-means clustering technique to find the best discrete ranges. And then we couple the discrete response variable with the center and variance of all the possible clusters of the continuous variable. It implies that every tour type is coupled with possible two or three clusters of the number of stops within that type of tour. This approach is more accurate than simply discretizing the continuous variable without taking into account the discrete response variable.

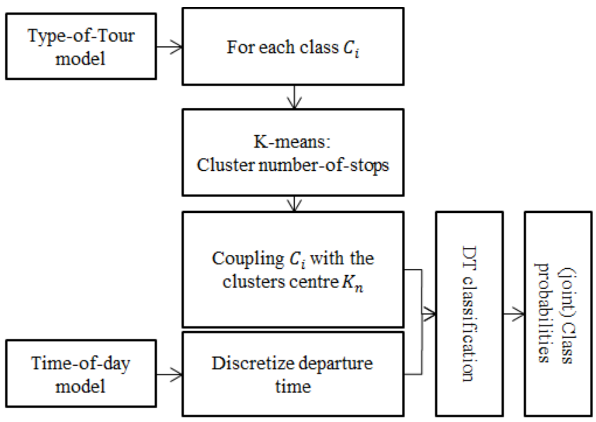

Figure 1: building blocks of the time-of-day and type-of- tour models

B. Decision tree

For this paper, we used Ctree from the Party package in R. For the details about the methodology we refer to [16] which describes the method implemented in this package. Here, we briefly highlight the main ideas. This algorithm has three steps:

- Take the independence test between any of the covariates and the response variable. Stop the algorithm if the null hypothesis cannot be rejected for all variables. Otherwise, choose the covariate X_j with the strongest correlation with the response variable Y.
- Choose the best subset A from X_j that can split X_j into two disjoint sets.
- For each of the splits, repeat steps 1 and two respectively.

This algorithm is based on permutation statistics tests and can work under a certain confidence interval without requiring any pruning or cross-validation [16]. The only parameters required include the required level of confidence (usually 95%) and the minimum number of cases in leaf nodes or the minimum number of weights that are required to split each node. This usually depends on the sample size of the training set.

IV. RESULTS AND DISCUSSION

In this section, we discuss the two decision tree models we developed with which we identify rules that can explain the structure of tours in freight activity of the agricultural products. Each resulting decision-tree is composed of nodes, which are numbered with positive integers from the root (e.g. top-node 1 in Figure 1) via intermediary nodes to the so-called leaf nodes (the bottom row of nodes: 3, 5, 6, 11, etc. in Figure 1). The first model describes the patterns in departure time of tours while the other model provides insight into the distinctions of the type of tours. Additionally, this latter model simultaneously predicts the average number of stops per tour type strategy. Both models are constructed using 7382 tours scheduled to transport agricultural products. We used a random sample of 80% of these data to estimate the model and the rest to test the (predictive) performance of the model. Figure 2 shows a schematic view of a tour structure and its components.

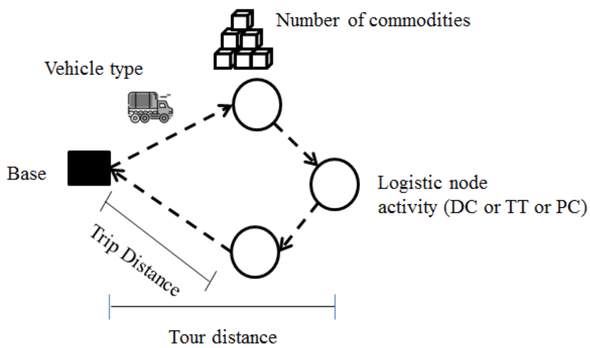

Figure 2:schematic of a tour structure and its components

Concerning all these components and dispatchers' perception of the congestion, we utilize the two developed models to identify the structure of tours by addressing the following five questions:

1. Under what circumstances are tours scheduled in off-peak hours and What type of tour strategies is planned to avoid congestion?
2. How does the activity of logistic nodes influence departure time and type of tour strategies?
3. How can tour/trip distance distinguish the structure of tours?
4. How can empty trips or the number of commodities affect the scheduling of tours?
5. How can vehicle type affect routing and scheduling of shipments?

A. Congestion related rules

In this section, we use both type-of-tour and time-of-day models to address question 1. We use all the explanatory variables listed in Table 1 as the input. Figures 3 and 4 are the estimated tree structure for the freight departure time schedules and type of tour strategies respectively. To understand the congestion avoidance strategies in this market, we look at nodes related to the congestion indicator of first and later intermediate stops. Following are significant rules that identify the structure of tours facing congested zones:

- Comparing leaf node 3 with leaf nodes 5 and 6 in figure 3 indicates that the chance of scheduling a tour before morning peak hours (before 6:00 AM) increases if visiting pick-up or delivery locations are located in congested zones.
- Dispatchers schedule Tours in the early morning or at night if the average trip distance between pick-up and delivery locations is more than 64 km (see Figure 3, nodes 4 and 5).
- Tours with 1 or 2 commodities have a high chance to depart at the evening peak period (Figure 3, node 16). This means that tours with a lower number of commodities usually do not avoid the peak period.
- In general, customers are served in a collection or distribution type of tours if they are located in congested zones (figure 4). On the other hand, planners schedule direct tours more often if there is no congestion.

B. Logistic Nodes specific rules

We considered three types of logistics nodes in this study: Distribution centers, transshipment terminal, and producer/consumers. However, the extracted rules related to the logistic node activity are very limited in both models. From model 1, we obtain three rules regarding to the departure time of tours visiting distribution centers (Figure 3, nodes: 14, 15, 23).

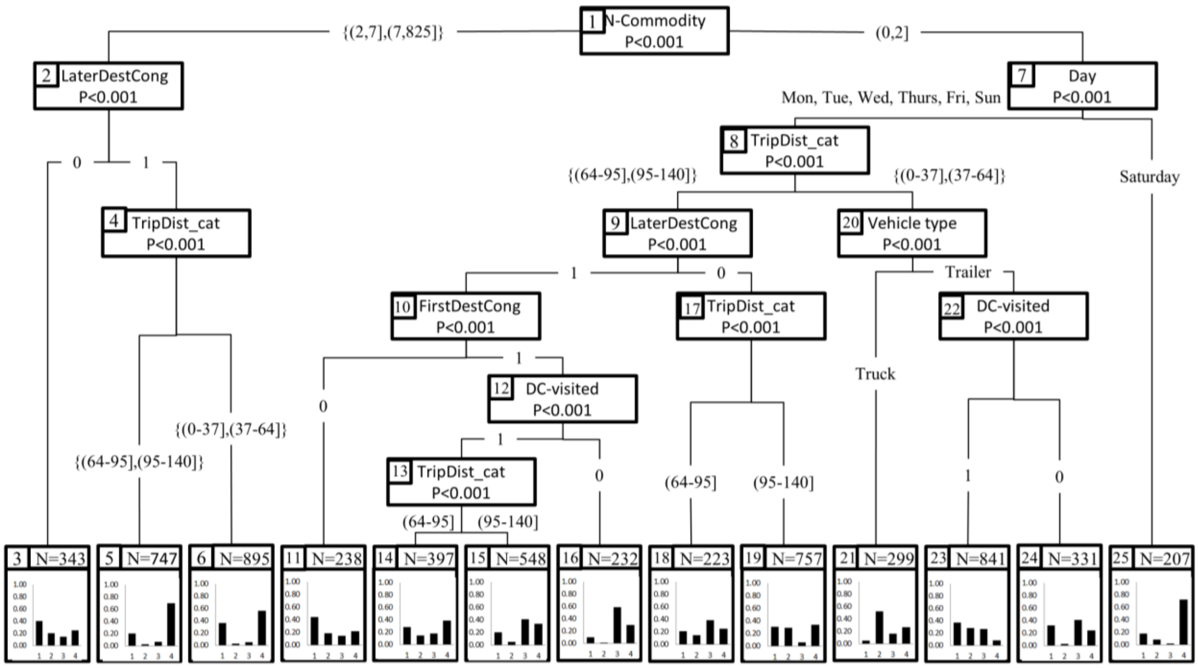

Figure 3:Estimated conditional inference tree for the scheduled departure time of tours

departure time

$$\rightarrow \begin{cases} (6:00-11:00], & \text{Tripdist} \leq 64\text{ km} \\ (15:00-20:00], & 64\text{ km} < \text{Tripdist} \leq 95\text{ km} \\ (20:00-6:00], & \text{Tripdist} > 95\text{ km} \end{cases}$$

Model 2 however does not show any significant rule for distribution centers. This is because most of the agricultural products in our data are distributed directly from producers and rarely go through other distribution centers. Therefore, the model does not suggest a significant distinction between tour type strategy of producer/consumer and distribution centers. It however identifies one significant rule for transshipment terminals:

- Tours visiting transshipment terminals are usually planned in a collection type of tour with four stops on average.

This rule implies that most of our tour data that go through transshipment terminals belong to the export of agricultural products. Third-party carriers pick-up commodities from several producers and deliver them to a transshipment terminal.

C. Trip/Tour distance

Tour and trip distance can be interpreted as transport costs. We obtained the following rule regarding tour and trip distance to address question 3.

- Figure 3 - Node 4 shows that the tours with higher trip distance (transport cost) are more likely to depart at night. The reason for this is that carriers usually should deliver commodities during working hours. Therefore, they depart at night/early morning not only to avoid congestion and reducing travel time costs but also to arrive during the working hours.

- We can see from Figure 4 – nodes 4 and 5 that the shorter the tour distance, the higher the number of stops in the distribution type of tour. In other words, planners may serve more local customers in one tour in a short distance and fewer customers in long distance.

D.Number of Commodities and empty trips

In this section, we address the question 4. Both models 1 and 2 indicate that number of commodities is one of the most important features to predict time-of-day and type of tour activities. The following rules are obtained from our DT models:

- Planners usually plan tours with more than 2 commodities in the early morning or night delivery (see Figure 3). One possible explanation for this rule is that the loading time is higher in this case and they must be scheduled in a way to avoid the peak periods and arrive on time.

- Neither model 1 nor model 2 identify any significant rule for empty trip (i.e. transporting empty pallets or empty containers)

- Given that the pickup and delivery locations are congested, Planners usually serve customers with :

  Tour type $\rightarrow$

  $$\begin{cases} \text{distribution}, & \text{number of commodity} > 7 \\ \text{collection}, & \text{number of commodity} \leq 7 \end{cases}$$

E. Vehicle type

Vehicle type is also one of the significant explanatory variables in both models. The following rules obtained from the models.

Mode 1 (Figure 3):

- Trucks have a higher chance of stating a tour between 11:00 AM and 3:00 PM.
- Trailers usually start a tour between 6:00 AM and 11:00 Am and also between 3:00 PM and 8:00 PM when there is no congestion.

Model 2 (Figure 4):

- For the vehicle type truck, the chance for the collection type of tour with four average stops and distribution type of tour with six stops is almost the same (see the node 13).
- Looking at the tours with one or two commodities, they are mostly scheduled in a collection type of tours. For this type of tour, if the vehicle type is truck, the average number of stops is in the higher category (8 stops on average, Node 15).
- However, if the vehicle type is a truck trailer, it depends on if a transshipment terminal is visited. in this case, the commodities are scheduled in a collection tour with a higher number of stops (8 stops on average, Node 17). Otherwise, they are more likely to belong to a collection tour with less number of stops (4 in average, Node 19 and 20).

V. MODEL PERFORMANCE AND EVALUATION

Table 2 gives information about the model parameters, predictive performance, and model fit. The accuracy of the models comes from the true positives, true negatives, false positives, and false negatives measures derived from the confusion matrix. However, to take into account the impact of imbalanced class distribution on accuracy, we report the F1-score, and one-vs-all (BAcc) accuracy. We also report the Kappa indicator which shows that how better the model is compared to predicting the class just by a random guess (23% in model 1). The goodness-of-fit e_{model} (0.47 for model 1) is equivalent to R^2 and is calculated based on the probabilistic theta proposed by [17]. e_{root} is the goodness of fit for a tree just on its root. The e_{incr} measures the improvement of the model compared to its root model. The parameter IS_i indicates the effect of each covariate on the prediction of the class *i* and the IS is an indication of the overall impact of each variable on the target variable. These coefficients are based on Chi-squared statistics between the frequency table predicted by the model for each variable and the expected frequency table assuming that there is no impact of that variable [17]. Besides the impact of the covariates, the direction of the impact is also interested. The MS_i is calculated based on the frequency of the class *i* under each level of the specific condition variable. This indicator is equal to 1 if the explanatory variable has a monotonically increasing impact on class *i* and is equal to -1 if it has a decreasing impact. If the indicator is between -1 and 1, the variable has a non-monotonous positive or negative impact on the class *i.* Note that this indicator is only meaningful for the ordinal or binary variables. For example. It does not mean anything for variable E2 which is the day of the week. All the coefficients in Tables 2 and their signs are also following the structure of the tree explained at the beginning of this section.

Model 2 shows a high accuracy (0.79), high balanced accuracy (0.91), and high kappa= 0.7 in predictions. It also has relatively high goodness-of-fit = 0.67 on test data. The parameter of the models shows the impact of each of the explanatory variables on the prediction of the target categories.

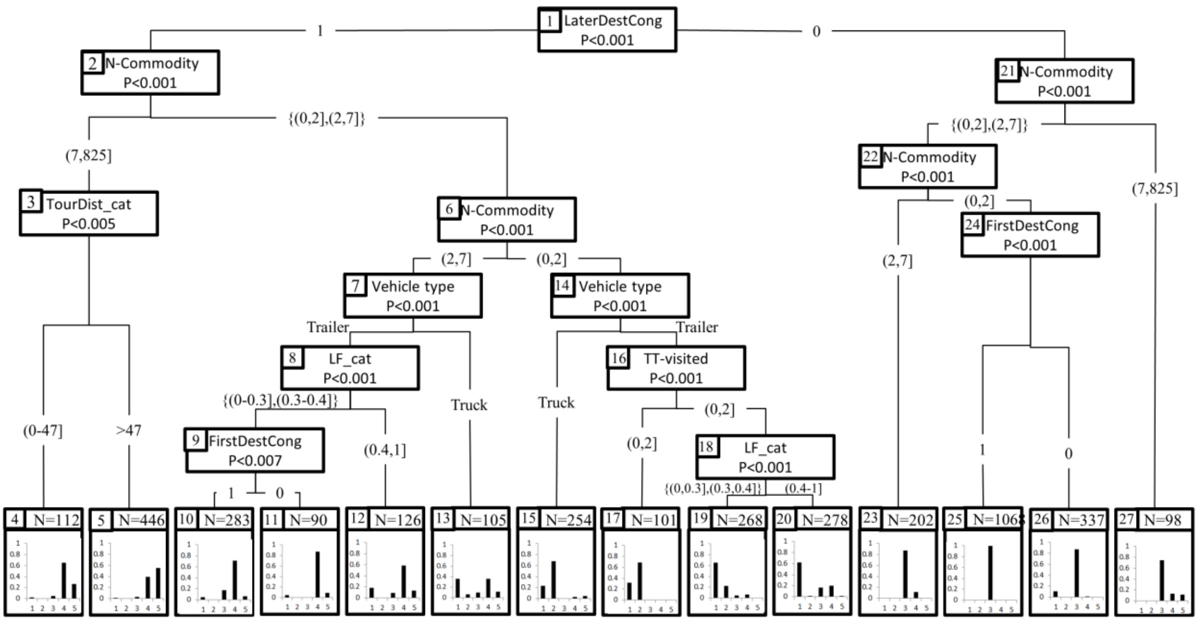

Figure 4. Estimated conditional inference tree for the joint model of the type-of-tour strategies and number of stops

TABLE II. MODELS PARAMETERS AND PERFORMANCE

Models		IS	IS_1	IS_2	IS_3	IS_4	IS_5	MS_1	MS_2	MS_3	MS_4	MS_5	P-value	performance
Model 1	E1	0.25	0.242	0.312	0.333	0.018	-	-1	-1	-1	0.11	-	0.0005	Acc=0.56
	E2	0.17	0.011	0.156	0.166	0.328	-	-1	-1	1	1	-	0.0005	BAcc=0.55
	E3	0.17	0.188	0.156	0.166	0.182	-	-1	1	1	1	-	0.0005	F1-score=0.39
	E5	0.16	0.217	0.156	0.166	0.098	-	1	1	1	1	-	0.0005	Kappa=0.23
	E6	0.10	0.176	0.048	0.062	0.215	-	1	1	1	1	-	0.0005	e=0.47
	E8	0.07	0.132	0.059	0.051	0.095	-	0.01	0.09	0.02	-0.02	-	0.0005	e_0=0.15
	E10	0.07	0.034	0.115	0.055	0.065	-	-0.12	-0.15	0.34	0.19	-	0.0005	e_{incr}=0.53
Model 2	E1	0.25	0.291	0.472	0.208	0.207	0.300	-1	-1	-1	0.13	1	0.0005	Acc=0.79
	E4	0.19	0.245	0.236	0.191	0.188	0.150	1	1	-1	1	1	0.0005	BAcc=0.91
	E5	0.05	0.077	0.098	0.036	0.053	0.025	1	1	1	1	1	0.0005	F1-score=0.71
	E6	0.13	0.016	0.035	0.189	0.142	0.064	-0.6	1	-0.4	-0.7	-0.4	0.0005	Kappa=0.7
	E7	0.17	0.243	0.015	0.185	0.177	0.123	-1	-1	-1	-1	-1	0.0005	e=0.67
	E9	0.13	0.110	0.044	0.138	0.179	0.141	1	-1	1	1	1	0.0005	e_0=0.46
	E10	0.07	0.018	0.101	0.052	0.055	0.198	-0.05	-0.2	-0.3	0.14	0.19	0.0005	e_{incr}=0.39

VI. CONCLUSION

In this paper, we proposed rule-based time-of-day and type of tour models which explain the general rules in freight transport of the agricultural product. The model's output explains how tours are scheduled in this market. Among all extracted rules from freight tour databases, The most outstanding findings are: (a) departure time of tours are sensitive to the congestion state of the pickup and delivery zones. (b) tours with high trip distances are more likely to depart in the early morning or night periods. (c) direct tours are usually planned for the uncongested zones. (d) in the agricultural industry, visiting a transshipment terminal happens more often in a collection type of tours. (e) the shorter the tour distance, the higher the number of stops in the distribution type of tour. The extension of this modeling can be useful for application in the activity-based disaggregate freight demand modeling and simulation.

ACKNOWLEDGMENT

We would like to thank NDW and CBS for providing us with traffic and truck diary data, respectively. This work was supported by the Netherlands Organization for Scientific Research (NWO), TKI Dialog, Commit2data, Port of Rotterdam, SmartPort, Portbase, TLN, Deltalinqs, Rijkswaterstaat, and TNO under project “ToGRIP-Grip on Freight Trips".

REFERENCES

[1] M. Khan, R. Machemehl, “Commercial vehicles time of day choice behavior in urban areas.”, Transportation Research Part A: Policy and Practice, 2017 Aug 1,102:68-83.

[2] Port-of-Rotterdam, “Continuing container growth pushes throughput at port of Rotterdam to a new high” [Press release], retrieved from https://www.portofrotterdam.com/en/news-and-press-releases/throughput2018, 2019.

[3] MA. Figliozzi “The impacts of congestion on commercial vehicle tour characteristics and costs”, Transportation research part E: logistics and transportation review. 2010 July 1;46(4):496-506.

[4] JD. Hunt, KJ. Stefan, “Tour-based microsimulation of urban commercial movements.”, Transportation Research Part B: Methodological, 2007 Nov 1,41(9):981-1013.

[5] A . Nuzzolo, A. Comi “Urban freight demand forecasting: a mixed quantity/delivery/vehicle-based model”. Transportation Research Part E: Logistics and Transportation Review. 2014 May 1;65:84-98.

[6] G. De Jong, M. Kouwenhoven, K. Ruijs, P. van Houwe, D. Borremans, “A time-period choice model for road freight transport in Flanders based on stated preference data”, Transportation Research Part E: Logistics and Transportation Review. 2016 Feb 1;86:20-31.

[7] AG. Vegelien, ER. Dugundji, “A Revealed Preference Time of Day Model for Departure Time of Delivery Trucks in the Netherlands.”, presented at the 21st International Conference on Intelligent Transportation Systems (ITSC), 2018 Nov 4 ,(pp. 1770-1774). IEEE.

[8] LD. Burns, RW Hall, DE. Blumenfeld, CF. Daganzo, “Distribution strategies that minimize transportation and inventory costs.”, Operations research, 1985 Jun,33(3):469-90.

[9] M. Ruan, JJ. Lin, K. Kawamura, “Modeling urban commercial vehicle daily tour chaining.”, Transportation Research Part E: Logistics and Transportation Review, 2012 Nov 1,48(6):1169-84.

[10] M. Khan, R. Machemehl. “Analyzing tour chaining patterns of urban commercial vehicles.”, Transportation Research Part A: Policy and Practice, 2017 Aug 1,102:84-97.

[11] M. de Bok, L.Tavasszy, “An empirical agent-based simulation system for urban goods transport (MASS-GT).”, Procedia computer science, 2018 Jan 1,130:126-33.

[12] RA. Mohammad, A. Nadi, M. de Bok, and LA. Tavatsszy, “Using data fusion to enrich truck diaries with firm databases”, unpublished.

[13] P. Christidis, JN. Rivas, “Measuring road congestion. Institute for Prospective and Technological Studies”, Joint Research Centre, Brussels. 2012 Apr.

[14] WY. Loh, “Fifty years of classification and regression trees.”, International Statistical Review. 2014 Dec;82(3):329-48.

[15] GV. Kass, “Significance testing in automatic interaction detection (AID)”. Journal of the Royal Statistical Society: Series C (Applied Statistics), 1975 Jun;24(2):178-89.

[16] T. Hothorn, K. Hornik, A. Zeileis,“Unbiased recursive partitioning: A conditional inference framework.”, Journal of Computational and Graphical Statistics, 2006 Sep 1,15(3):651-74.

[17] T. Arentze, H. Timmermans, “Measuring Impacts of Condition Variables in Rule-Based Models of Space-Time Choice Behavior: Method and Empirical Illustration.”, Geographical Analysis, 2003 Jan,35(1):24-45.