Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 00 (2015) 000–000

Procedia
Computer Science

www.elsevier.com/locate/procedia

6th International Conference on Ambient Systems, Networks and Technologies, ANT 2015

# A data mining based method for route and freight estimation

Shoaib Bakhtyar[a,*], Johan Holmgren[a,b]

[a]*Department of Computer Science and Engineering, Blekinge Institute of Technology, Karlskrona SE-371 79, Sweden*
[b]*Department of Computer Science, Malmö University, Malmö SE-205 06, Sweden*

**Abstract**

We present a method, which makes use of historical vehicle data and current vehicle observations in order to estimate 1) the route a vehicle has used and 2) the freight the vehicle carried along the estimated route. The method includes a learning phase and an estimation phase. In the learning phase, historical data about the movement of a vehicle and of the consignments allocated to the vehicle are used in order to build estimation models: one for route choice and one for freight allocation. In the estimation phase, the generated estimation models are used together with a sequence of observed positions for the vehicle as input in order to generate route and freight estimates. We have partly evaluated our method in an experimental study involving a medium-size Swedish transport operator. The results of the study indicate that supervised learning, in particular the algorithm *Naive Bayes Multinomial Updatable*, shows good route estimation performance even when significant amount of information about where the vehicle has traveled is missing. For the freight estimation, we used a method based on averaging the consignments on the historical known trips for the estimated route. We argue that the proposed method might contribute to building improved knowledge, e.g., in national road administrations, on the movement of trucks and freight.

Peer-review under responsibility of the Conference Program Chairs.

*Keywords:*
Route estimation; freight estimation; supervised learning; Naive Bayes multinomial;

## 1. Introduction

It is commonly agreed that road freight transport contributes both positively (e.g., concerning economy and social welfare) and negatively (e.g., pollution, congestion, noise, and road accidents) to a society. An important challenge of today is to achieve an efficient and sustainable transport system, where the positive effects are maximized at the same time as the negative effects on the environment are minimized. A possible means to succeed with this challenge is to apply different types of policy instruments, such as taxes, fees, and infrastructure investments. However, significant changes to the transport system should not be done without first acquiring accurate information on the current system.

Large amounts of resources are spent yearly on collecting traffic data, e.g., in order to estimate the transport volumes and travel times on roads in congested areas. Still, there is often a large gap in the knowledge concerning the transport system, in particular regarding the movement of individual vehicles and of freight. We argue that this

* Corresponding author. Tel.: +46-734-223921.
*E-mail address:* shoaib.bakhtyar@bth.se

1877-0509 
Peer-review under responsibility of the Conference Program Chairs.

gap is partly caused by the limitations of the technologies that are used for measurement. The dominating method in use, at least in the country of our study, is pneumatic tubes, which allows counting vehicles, as well as estimating vehicle types and traveling speeds. However, the method does not allow identifying vehicles, and the collected data can therefore not be used to determine how individual vehicles have traveled in the network. In addition, the data collection methods of today do not allow estimating what the observed vehicles carry.

The purpose of the current paper is to contribute to the development of methodology for building knowledge on the movement of trucks and freight in a transport network. In particular, we contribute a method for estimating the route a vehicle used and the freight carried by that particular vehicle, based on information about where the vehicle usually travels and the freight it usually carries. The method includes a learning phase and an estimation phase, where the learning phase uses historical data about the movement of a vehicle and the consignments allocated to the vehicle in order to generate estimation models for the route choice and consignment allocation. The estimation phase makes use of the generated estimation models to estimate the route and the carried freight, using only a sequence of vehicle positions as input, i.e., positions where the vehicle have been observed. Our work is based on the assumption that the vehicles traveling in a network, to a sufficiently large extent, can be identified along the roads, e.g., using road-side cameras that operate together with automatic number plate recognition software or vehicle-to-vehicle communication (e.g., using transponder technology). We believe that our research might be of interest for actors, mainly public road administrations, who are able to observe (and identify) a vehicle at different locations in a network, or who in other ways are able to collect information about vehicles positions. They also need historic data about the routes the vehicle usually uses and the freight it usually carries. Perhaps, these types of information could be collected using extended commodity flow surveys, where also route data is collected. Commodity flow surveys are conducted on a regular basis in some countries, e.g., in the US. In particular, the presented method can be useful in situations where the interest is on identifying the freight types and volumes and routes used by vehicles, in order to support the public policy making, e.g., concerning road taxation.

An enabler for methods like the one presented in the current paper is the fast technological development that we are currently experiencing; a development that leads to new opportunities for collecting traffic data and converting the data into useful traffic information. In particular, it leads to increasing amounts of digitally stored data about consignments and vehicle movements, improved technologies for identifying vehicles, and improved methods for analyzing and finding patterns in large collections of data.

We partly evaluated the proposed method in collaboration with a medium-size Swedish road freight transport operator, who provided data about the movement (GPS trajectories) and allocated consignments (order data) for one vehicle for a period of one month. In particular, we used the received data in order to build and evaluate route and freight estimation components for the suggested method.

The rest of the paper is organized in the following way. In Section 2, we outline related work. In Section 3, we present our method for route and freight estimation. In Section 4, we describe our experimental evaluation of the suggested method, and in Section 5, we conclude the paper with a discussion and some pointers to future work.

## 2. Related Work

There exist different types of freight transport analysis models for estimating freight transport under different conditions, and for supporting the design of the transport network. The dominating type of method is the so-called 4-step freight transport analysis models[1], e.g., Samgods[2] and TRANS-TOOLS[3]. However, it is often difficult to guarantee the quality of the vehicle and freight flow estimates that are generated using this type of models, e.g., since the quality of the input data used in the model might not be sufficiently good, and the models typically make several estimates, e.g., of vehicle types, traffic modes, transport chains, and routes, before arriving at the final vehicle and freight estimates. There also exist other types of models for the same purpose; see, De Jong et al.[1], Chow et al.[4], and Holmgren et al.[5] for overviews.

Another field of related research is map-matching, where the purpose is to translate a sequence of vehicle locations into the correct route in a traffic network (see, e.g., Quddus et al.[6] for an overview). Brakatsoulas et al.[7], for example, propose and evaluate three map-matching algorithms that consider the trajectory nature of the GPS data and which have the ability to map a trajectory to a road network.

The most closely related work is by Froehlich et al.[8], Simmons et al.[9], and Bakhtyar et al.[10]. Froehlich et al. present different algorithms for predicting a driver's route based on historical trip data for the same driver. They test the algorithms on data acquired for 250 drivers over a period of 15 days. In order to predict a driver's route, they compare the first part of a drivers current trip with the previously observed trips for the same driver. Simmons et al.[9] propose an approach for predicting a driver's intended destination and route by applying a hidden markov model on observations of the driver's driving habits. Through the use of online GPS observations for a driver, they manage to predict the destination and route for that particular driver using the proposed model. They claim that their model is able to achieve 98% accuracy in most cases. The work by Bakhtyar et al.[10] concerns the estimation of freight types between terminals (i.e., the origin and destinations of consignments under transport). In particular, they present a comparison of classification algorithms for the considered problem. In Table 1, we present the difference between the work in this paper and the most closely related work.

Table 1. Our research compared to other closely related work.

| | Estimation type | Transport type | Intended user(s) |
|---|---|---|---|
| Froehlich et al.[8] | Route estimation | Passenger | Driver |
| Simmons et al.[9] | Route estimation | Unknown | Driver |
| Bakhtyar et al.[10] | Freight estimation | Freight | Public authorities |
| Our study | Route and Freight estimation | Freight | Public authorities |

The main difference between the focus of this paper and the related research discussed above is that we suggest building estimation (prediction) models for both freight and route. In particular, we propose a method that explicitly utilize on historical data (trips and freight data) in order to translate a set of vehicle positions into route and freight estimates. Another key difference between our work and the closely related work is the intended user; our aim is to support the decision making by public authorities.

## 3. A method for route and freight estimation

In this section, we present our method for building and using estimation models for the route used by a vehicle and of the freight carried on that particular route. The method includes a learning phase and an estimation phase. The learning phase uses historical data about the movement of the vehicle and the consignments allocated to a vehicle in order to build the estimations models: one model for the route choice and one model for the freight allocation. In the estimation phase, the estimation models are then used in order to estimate the route and freight using only a sequence of vehicle positions as input. The proposed route and freight estimation method is illustrated in Fig. 1.

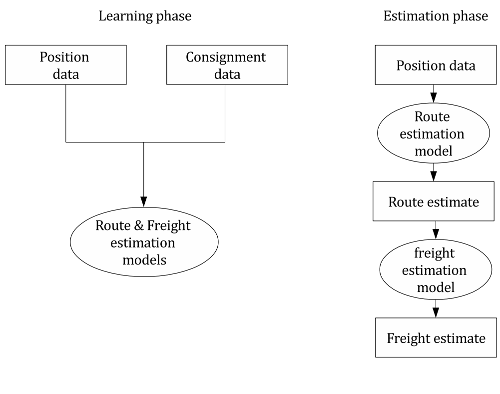


Fig. 1. Our method for route and freight estimation.

## 4. Evaluation of the estimation method

We partly evaluated our route and freight estimation method in a study involving a medium-size Swedish road freight transport operator, who provided GPS data (from their fleet management system) and consignment data (from their customer order system) for one of their trucks during a period of one month. We verified the correctness of the received data by comparing the waybills for the truck with the order data for the same truck. In addition, we installed monitoring equipment in the truck, and compared the collected data with the received GPS data. From the collected data we identified 1) the outbound trips in the GPS data and 2) which of the consignments that belonged to each of the outbound trips. We here define an outbound trip as a sequence of links, where the first link starts and the last link ends in the home terminal of the truck; each time the vehicle visits its home terminal, it initiates a new outbound trip. In the data processing phase, we grouped the locations of the pickup and deliveries in the consignment data so that all of the pickups and deliveries located in the same town were represented by one "terminal", where freight may be picked-up or delivered. In total, we identified 18 bidirectional links (L1,...,L18) that connect the 18 terminals (N1,...,N18). In Fig. 2, we illustrate the considered network. In total, the studied truck made 11 outbound trips along the 7 routes (described below as sequences of nodes). For future reference, we present the trips that belong to each of the routes, where trip 1 is oldest, trip 2 is second oldest, etc.

**R1:** (trip 2) N1-N2-N3-N4-N5-N4-N3-N2-N1

**R2:** (trip 3, 6, and 8) N1-N7-N8-N9-N10-N11-N13-N14-N15-N16-N17-N18-N17-N16-N15-N14-N13-N11-N10-N9-N8-N7-N1

**R3:** (trip 5 and 7) N1-N7-N8-N9-N10-N11-N13-N14-N15-N16-N17-N17-N16-N15-N14-N13-N11-N10-N9-N8-N7-N1

**R4:** (trip 1 and 4) N1-N7-N8-N9-N8-N7-N1

**R5:** (trip 9) N1-N7-N8-N9-N10-N11-N13-N14-N15-N16-N17-N16-N15-N14-N13-N1

**R6:** (trip 10) N1-N2-N3-N4-N3-N2-N1

**R7:** (trip 11) N1-N2-N3-N4-N7-N8-N9-N10-N11-N12-N11-N10-N9-N8-N7-N4-N3-N2-N1

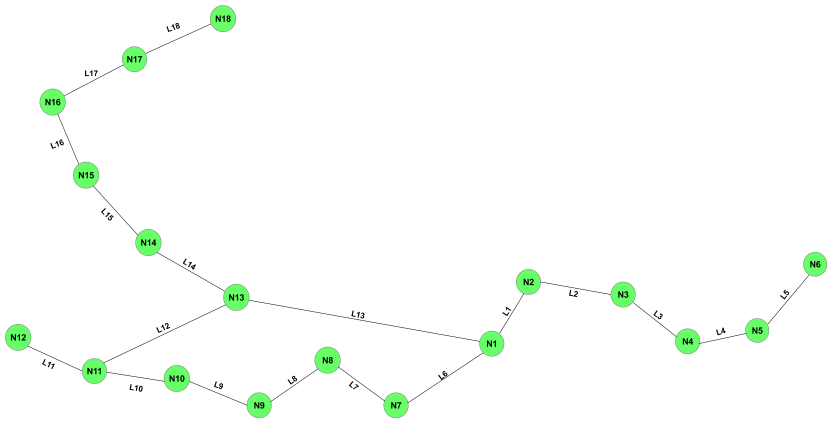


Fig. 2. Network used by the studied truck.

We grouped the transported freight into two categories: 1) food and consumption products and 2) non-food items. The food and consumptions products category includes fish, meat, vegetables, etc., and the non-food freight consists of several types of products, e.g., transport equipment, furniture, and mail. The reason for using only two categories is that most of the studied consignments concern various types of food items and consumption products that could not be distinguished from each other. In addition, there was such a large variety among the other product types that it would not be possible to identify any patterns unless they were aggregated into one category.

It should be noted here that we had order data for a period of 25 days, whereas the GPS data covered 33 days. For this reason, we did not have any order data for R5, R6, and R7. However, in order to evaluate the route estimation phase of our model, we chose to keep R5, R6, and R7. In total, we had order data for six of the identified 11 trips, that is, trips 2, 3, 5, 6, 7, and 8.

### 4.1. Route estimation

As mentioned above, the route estimation should be made using only a set of vehicle positions, representing the links where the vehicle has been observed, as input. In particular, it should be able to work with high accuracy even if the vehicle has been observed only on a subset of the links along a route. It can be argued that each of the routes in our dataset defines a class, and since the route estimation concerns estimating the most probable route (or class) from a set of vehicle observations (i.e., the links where the vehicle has been observed), we used supervised learning for the route estimation. In particular, we used the open source *Waikato Environment for Knowledge Analysis* (Weka)[11].

We considered different ways of representing the routes in the learning algorithms; we finally chose a route representation that explicitly specifies for each of the 18 links in the extracted network whether or not it is included in the route. Our representation of the identified routes is provided in Table 2, where each row represents a route, the first column is the route label (i.e., the class), and each of column 2-19 specifies if the corresponding link is in the route (value 1) or not (value 0). In order to evaluate our route estimation approach, we used the identified seven routes, as specified in Table 2, as the training set. The test set consists of all the routes that can be generated by removing one or more link from the seven routes, as will be detailed below.

Table 2. Route representation in the supervised learning algorithms.

| | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 | L9 | L10 | L11 | L12 | L13 | L14 | L15 | L16 | L17 | L18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| R2 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| R3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| R4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| R5 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| R6 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| R7 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

As mentioned above, the estimation (prediction) corresponds to identifying the correct route (or label) from a, typically, incomplete description of a route. An incomplete route description corresponds to the fact that the vehicle has been observed at one or more (but not at all) locations along a route. In other words, there are at least one link along the route for which there exist no observation. An incomplete route description corresponds to a row in Table 2, where one or more of the ones have been exchanged with zeros. In order to reduce the size of the test set, we did not include the links used by the vehicle when returning to the home terminal. However, in the freight estimation part, we still consider those links.

By requiring that an incomplete route description should contain at least one link, there are 5666 possible incomplete route descriptions for our seven routes. These 5666 instances constitute our test instances. For a route with $n$ links, there exist $\binom{n}{k}$ ways to remove $k$ links (or changing $k$ ones to zeros). For example, for R1, which has 5 links, there are $\binom{5}{1} = \binom{5}{4} = 5$ ways to remove 1 or 4 links and $\binom{5}{2} = \binom{5}{3} = 10$ ways to remove 2 or 3 links.

We initially investigated each of the supervised learning algorithms provided by Weka, by classifying all of 5666 test instances. The algorithm *Naive Bayes Multinomial*, in particular the *Naive Bayes Multinomial Updatable*, performed best on the considered dataset with 61.4% correctly classified instances. We therefore chose this algorithm to study in more detail. The *Naive Bayes Multinomial Updatable* is an updated version of the classical *Naive Bayes Multinomial* algorithm, which is mainly used for document classification[12]. The *Naive Bayes Multinomial* views a document as a set of words by considering the word counts (i.e., words frequency). The *Naive Bayes Multinomial Updatable* also considers patterns in the words frequency in the document[13, 14]. Although 61.4% seems rather low, it should be noted that the test set included the worst-case route descriptions, e.g., it includes for R5 (with 11 links) 11 instances with only one link, 55 instances with two links, and 165 instances with three links.

We divided our test set into 10 subsets: one subset with all incomplete route descriptions containing one link, one subset with all incomplete route descriptions containing two links, etc. From the confusion matrices generated for our 10 test subsets, we observed that the higher the number of links in the incomplete route descriptions, the higher the share of correctly classified instances will typically be. This can also be seen in Fig. 3, where we show for each of the routes the share of correctly classified test instances with one link kept, two links kept, etc., for *Naive Bayes Multinomial Updatable*.

The routes R4 and R6, which had the least number of links (i.e., 3 links), were always correctly classified. Therefore, we chose not to display them in Fig. 3. The routes R1, R3, and R7 were always correctly classified

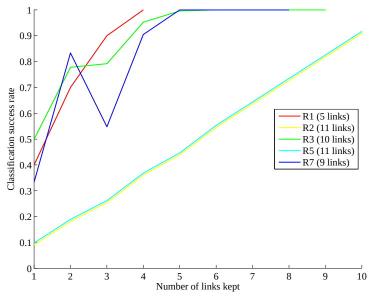


Fig. 3. For route R1, R2, R3, R5, and R7, the share of correctly classified test instances with one link kept, two links kept, etc.

for all incomplete route descriptions with at least 4 and 5 links respectively. The routes R2 and R5 (having 11 links each) showed an identical classification success pattern. The reason for this could be that these routes differ with only one link: R2 includes L18 but not L13 while R5 includes L13 but not L18. Both R2 and R5 can be classified correctly up to 91% with 10 links present. The reason for not observing a higher classification success rate than 91% is that R2 and R5 becomes identical to R3 when removing L18 and L13, respectively. In Fig. 3, it can be seen that the classification rate for R7 drops when the number of links are increased from 2 to 3. This is because 45% of the instances of R7 were incorrectly classified as instances of R4 and R6. The reason for this behavior is that there is a good match with R4 and R6 when removing all but three links from the description of R7. For example, all 18 incomplete representations of R7 with three links, where at least one of L1, L2, and L3 and none of L6, L7, and L8 are represented, will be misclassified as R6. Altogether, we argue that the presented results show that the supervised learning approach, in particular the *Naive Bayes Multinomial Updatable* algorithm, has potential to perform rather well for route estimation.

### 4.2. Freight estimation

The purpose of the freight estimation part of our estimation method is to estimate the freight types and corresponding weights carried by the vehicle along the estimated route. We chose to estimate the freight as the average of the historical consignments for the outbound trips for the selected route. For example, if there are two outbound trips for a route with the same freight types, however in different amounts (lets say $x$ and y kg), that are carried between the same origin and destination, we estimate the amount of freight carried by the vehicle as $(x+y)/2$ kg between the two locations. We used this approach as our freight estimation, since it is a straightforward way to estimate the carried freight, and it takes into account all the information about the consignments previously carried by the studied vehicle along a particular route.

We identified the consignments for a given trip in the order data set, by identifying those consignments that were loaded the same day as the outbound trip, and had loading and/or unloading locations that matched the sequence of terminals in the outbound trip. For some of the consignments, we were only able to identify a loading or an unloading location. We assumed that these consignments were loaded or unloaded, respectively, in the home terminal of the vehicle, i.e., node N1. We further assumed that some other trucks completed these orders. For the same reason as we had consignments assigned to the studied truck, which were partially completed by other trucks, there could be consignments assigned to other trucks that were partially completed by the considered truck. Therefore, we realized that there is risk that we underestimated the freight loaded on the truck. This was later on confirmed by the transport operator involved in our study.

In Fig. 4a, 4b, 5a, and 5b, we present, for the purpose of illustration, our freight estimates based on the identified trips along R2 (trips 3, 6, and 8) and R3 (trips 5 and 7). In the figures, the links are presented on the x-axes and the amounts of freight (in kilograms) carried on each of the links are presented on the y-axes. Fig. 4a and Fig. 5a represent the amount of food and consumption products carried along R2 and R3, respectively. Fig. 4b and Fig. 5b represent the amount of non-food items on R2 and R3, respectively. It should be noted that we do not present freight estimates for R1, because we observed (and the company confirmed) that the order data for trip 1 was not representative for the studied truck.


(a) (b)

Fig. 4. Freight estimate diagrams for (a) food freight, and (b) non-food freight for R2.


(a) (b)

Fig. 5. Freight estimate diagrams for (a) food freight, and (b) non-food freight for R3.

In particular, the increases and decreases of the carried freight, which can be observed in the diagrams, clearly show how the truck is estimated to have carried the freight along the two routes. For example, for trip 3 and 6, we observed from Fig. 4a that food freight was carried on L6, L7, L8, L9, L10, and L12 before the same type of freight is unloaded, i.e., on the nodes connected by L14, L15, L16, L17, and L18. It can be observed that on L18 and L18', the freight weight decreases sharply. The reason for a steep decrease or increase on a particular link is due to the unloading or loading of large freight volumes (i.e., weight) before the vehicle enters the link. The same pattern can be seen on L18' and L17', where there is a steep increase in the freight weight.

For evaluation purposes, we presented the results from the freight estimation to the manager of the transport company included in the study. He confirmed that the generated freight estimates for R2 and R3 correspond well with the typical pattern of the studied vehicle. He explained that the reason for the high difference in the non-food freight amount on trip 8 compared to trips 3 and 6 on route R2 (see, Fig. 4b) is due to a temporary load of 20 tons.

## 5. Concluding remarks

We have presented a method for route and freight estimation, which includes a learning phase and an estimation phase. The purpose of the learning phase is to build estimation models based on historical vehicle data; one model for route choice and one model for freight allocation. The purpose of the estimation phase is to estimate what freight (types and weights) was carried by the vehicle along an estimated route. We partly evaluated the proposed method in a study involving a medium-size Swedish road freight transport operator. We conclude that the method has potential to provide good route and freight estimations, as long as there exist sufficient amounts of historical vehicle movement and consignment data, as well as current vehicle observations. In addition, we conclude that more detailed data (typically found in fleet management and ordering systems) is needed when building the prediction model, whereas only positions (typically sparse) along the road is needed in order to estimate routes. We used supervised learning for the route estimation, i.e., the *Naive Bayes Multinomial Updatable* algorithm, which performed rather well for route estimation. For freight estimation, we estimated the freight carried by the vehicle along a particular route as the average of the consignments that previously have been observed on the outbound trips for the route.

A possible limitation of the suggested method, which estimates the route and freight in two sequential steps, is that incorrect route estimations might lead to incorrect freight estimations. However, our evaluation indicate that it is possible to achieve a rather accurate route classification using supervised learning, unless very little information is known about the movement of the vehicle. As the method operates in two steps, i.e., first the route is estimated, and then the freight is estimated, we find it reasonable to assume that the performance of the method to a large extent depends on the choice of model components for the route and freight estimation, respectively.

Even though there is a risk of transferring errors between the route and freight estimation models, we argue that there are also desirable features of the approach. Mainly, a two-step (sequential) approach, where the interface between the steps is clearly defined, enables to replace only one of the estimation models (either route or freight) and keep the other one. The possibility to independently choose models for the two steps is an important reason that we suggest using an approach where route and freight is estimated sequentially.

An interesting direction for future work is to estimate freight types and volumes on routes for which there exist no order data. Additionally, the proposed method needs to be evaluated in more detail, in particular, the freight estimation part of the approach, which we only evaluated through expert opinion.

In addition, it is relevant to study whether it is best to create one estimation model for each vehicle of a company, or if it is better to cluster vehicles so that prediction models represent multiple vehicles. It is also relevant to study the possibility to include, in the estimation models, that consignments may be transferred between vehicles, which in our study led to an underestimation on carried freight for some links.

## Acknowledgements

The presented research was conducted within the project Arena3, which was led by NetPort Karlshamn, and funded by the Regional Council of Blekinge, the Swedish Governmental Agency for Innovation Systems, the Swedish Transport Administration, and the Swedish Transport Agency. In addition to the above, we would like to thank the Swedish road freight transport operator who made our research possible by providing consignment and vehicle movement data, and by allowing us to monitor one of their trucks. We would also like to thank the Swedish National Postgraduate School on Intelligent Transport Systems (NFITS) for supporting this research.

## References

1. De Jong, G., Gunn, H., Walker, W.. National and international freight transport models: An overview and ideas for future development. *Transport Reviews* 2004;**24**(1):103–124.
2. Swahn, H.. The swedish national model system for goods transport-samgods: A brief introductory overview. *SAMPLAN RAPPORT* 2001;(2001: 1).
3. Rich, J., Bröcker, J., Overgård, C.H., Korzenewych, A., Nielsen, O.A., Vuk, G.. Report on scenario, traffic forecast and analysis of traffic on the ten-t, taking into consideration the external dimension of the union: Trans-tools version 2; model and data improvements. Tech. Rep.; 2009.
4. Chow, J.Y., Yang, C.H., Regan, A.C.. State-of-the art of freight forecast modeling: lessons learned and the road ahead. *Transportation* 2010;**37**(6):1011–1030.
5. Holmgren, J., Davidsson, P., Persson, J.A., Ramstedt, L.. Tapas: A multi-agent-based model for simulation of transport chains. *Simulation Modelling Practice and Theory* 2012;**23**:1–18.
6. Quddus, M.A., Ochieng, W.Y., Noland, R.B.. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C: Emerging Technologies* 2007;**15**(5):312–328.
7. Brakatsoulas, S., Pfoser, D., Salas, R., Wenk, C.. On map-matching vehicle tracking data. In: *Proceedings of the 31st international conference on Very large data bases*. VLDB Endowment; 2005, p. 853–864.
8. Froehlich, J., Krumm, J.. Route prediction from trip observations. Tech. Rep.; SAE Technical Paper; 2008.
9. Simmons, R., Browning, B., Zhang, Y., Sadekar, V.. Learning to predict driver route and destination intent. In: *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*. IEEE; 2006, p. 127–132.
10. Bakhtyar, S., Henesey, L.. Freight transport prediction using electronic waybills and machine learning. In: *Informative and Cybernetics for Computational Social Systems (ICCSS), 2014 International Conference on*. IEEE; 2014, p. 128–133.
11. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.. The weka data mining software: an update. *ACM SIGKDD explorations newsletter* 2009;**11**(1):10–18.
12. Witten, I.H., Frank, E.. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann; 2005.
13. McCallum, A., Nigam, K., et al. A comparison of event models for naive bayes text classification. In: *AAAI-98 workshop on learning for text categorization*; vol. 752. Citeseer; 1998, p. 41–48.
14. Razavi, A.H., Inkpen, D., Uritsky, S., Matwin, S.. Offensive language detection using multi-level classification. In: *Advances in Artificial Intelligence*. Springer; 2010, p. 16–27.