Forecasting Travel Times with Space Partitioning Methods

Jhonny Pincay[1][a], Alvin Oti Mensah[2], Edy Portmann[1] and Luis Terán[1,3][b]
[1] *Human-IST Institute, University of Fribourg, Boulevard de Pérolles 90, Fribourg, Switzerland*
[2] *University of Bern, Hochschulstrasse 6, Bern, Switzerland*
[3] *Universidad de las Fuerzas Armadas ESPE, Av. General Rumiñahui S/N, Sangolquí, Ecuador*

Keywords: Travel Time, Spatio-temporal Data, Transportation, Smart Logistics, Geohash, Geogrid.

Abstract: Roads and streets are more and more crowded. For delivery companies that use road transportation, this is a concerning issue as longer times spent on roads mean higher operational costs and less customer satisfaction. Nevertheless, the data captured during operation hours of their vehicles can be leveraged to address such issues. This, however, is not a straightforward task given the possible low number of vehicles covering one route and the complexities introduced by the delivery business nature. The present research work proposes an approach to forecast travel time through the use of probe data from logistic vehicles and simple mathematical models. The delivery operations of five months of a vehicle from the Swiss Post, the national postal service company of Switzerland, were studied in a segment-to-segment manner, following a four-step method. Moreover, the results of the forecasting were evaluated calculating the mean absolute percentage error and mean absolute error metrics. The results obtained indicate that is possible to achieve a considerable forecasting accuracy without the deployment of a large number of vehicles or the implementation of complex algorithms.

1 INTRODUCTION

The number of vehicles on roads and streets has massively increased over the decades, which translates into more frequent traffic congestion. This has led to the need for people to plan their journeys pre-trip and en-route. Moreover, for companies, considering vehicle transit and their impact on their supply chain has become vital to keep their operational costs as low as possible. Factors as the aforementioned have sparked a growing interest in traffic modeling and forecasting of travel times. The technological advent and the data availability of recent years have also eased the development of systems that provide travel time information to commuters. Common data sources used for such systems include sensors (e.g., point and loop detectors), studies on-site, and global positioning systems (GPS) equipped in vehicles circulating on roads (Mori et al., 2015; Zhou et al., 2012).

Traffic and travel time modeling using GPS data has gained considerable attention as it is one of the less costly methods. A growing number of studies are devoted to developing such models employing data that come from fitted vehicles used for traffic data gathering, taxis, public transportation, among others. Common methods to process such data are machine learning and deep learning-based, which however offer good results, require vast amounts of data and computational resources (Pu et al., 2009; Zhou et al., 2012; Yuan et al., 2010).

On the other hand, little to no attention is given to traffic data from logistics vehicles, given the complexities introduced as a consequence of business operations. Road and speed restrictions, multiple delivery stops, and waiting on customers are some of the events recorded in delivery probe data which need to be properly handled when building travel time models. Another issue is the low sample rate as logistic companies might have unique vehicles covering certain routes, yet it is also possible that this vehicle circulates through the same route every day and thus, large amounts of data are produced. If the data collected by the logistic vehicles is properly studied, important insights can be obtained which could be used to draw insights from enterprise supply chains for strategic planning.

This research project proposes a novel approach for network-wide travel time estimation, in sight of the constraints previously highlighted. In this con-

[a] https://orcid.org/0000-0003-2045-8820
[b] https://orcid.org/0000-0002-0503-511X

Pincay, J., Mensah, A., Portmann, E. and Terán, L.
Forecasting Travel Times with Space Partitioning Methods.
DOI: 10.5220/0009324601510159
In *Proceedings of the 6th International Conference on Geographical Information Systems Theory, Applications and Management (GISTAM 2020)*, pages 151-159
ISBN: 978-989-758-425-1

text, a data-driven approach for travel time estimation, using data from a company's probe-vehicles and geospatial indexing is proposed. As a result, it is expected to define a straightforward method that allows forecasting travel times with acceptable levels of accuracy. For this effect, an artifact was developed following the principles of the design science methodology.

This article is structured as follows: Section 2 presents the theoretical background on which this research work is grounded. Then, the methods used in this study are described in Section 3. Results are presented in Section 4. Section 5 finalizes the article with a summary and concluding remarks.

2 THEORETICAL BACKGROUND

This section presents the concepts used in this research work. Some previous research efforts that attempted to achieve similar goals are also examined.

2.1 Traffic and Travel Time Estimation

Lin et al. (2005) defined the main components of a road traffic environment as humans, vehicles, and facilities (e.g., roads and signaling). Humans and vehicles constitute traffic demand, whereas the facilities provide the supply. According to this notion, travel time is dependent on the dynamism and interactions between the demand and supply and the conditions affecting any of them (e.g., road nature and weather).

Furthermore, road traffic can be classified into two states: (i) congested/jam (ii) uncongested/free flow (Treiber and Kesting, 2013). There is a set of measurable traffic characteristics or variables, capable of describing the traffic in any of these two states. These variables are referred to as *traffic state variables*. The fundamental traffic variables include flow, vehicle density, and speed.

Aside from these three variables, there are other equally important traffic variables such as the *travel time* (Nanthawichit et al., 2003; Van Lint and Van Hinsbergen, 2012). The majority of methods dealing with traffic and travel time analysis depend on the full availability of the aforementioned variables. Data can be collected using externally localized traffic measuring instruments, which record a comprehensive state of the traffic conditions within their coverage range (Treiber and Kesting, 2013). Data captured from these stationary devices is known as *trajectory data*. Even though this approach allows having a full picture of traffic at any given point in time, the number of devices that need to be deployed is rather high and therefore expensive (Ruppe et al., 2012; Yoon et al., 2007).

Nevertheless, some methods allow working with partially observed or incomplete data. These methods are known as traffic state estimation (TSE). According to Seo et al. (2017), TSE is the process of deducting traffic state variables on road segments (portion of a road) using partially observed data. Such methods can be model-driven, data-driven or streaming-data-driven. The general approach of TSE methods in performing traffic data analysis is characterized by D'Andrea and Marcelloni (2017) and Wang et al. (2013) into three phases:

- **Segmentation.** Divide roads into finer spatial and/or temporal units (segments).
- **Annotation.** Annotate segments with an expected behavior (e.g., vehicle density, travel time).
- **Estimation.** Inference with with respect to the expected behavior for each segment

In TSE methods where estimations of travel times are performed at finer spatio-temporal resolution, travel time is defined as the amount of time taken to traverse a unit space of a road segment, usually measured in minutes per kilometer (min/km) (Seo et al., 2017). At a micro-scale, travel time is calculated for individual vehicles given their respective entry and exit times in a segment. The travel time, therefore, is calculated using Equation 1.

$$TT^i = \frac{T^i_{out} - T^i_{in}}{D}\ min/km \tag{1}$$

where, TT^i is the microscopic-scale travel time for a vehicle i. T_{out} is the timestamp at which the vehicle exits the segment. T_{in} is the timestamp at which the vehicle entered the segment. D is the length of the segment. Aggregating individual travel times in a segment estimates the travel time for a segment at a macro-scale. A macro-scale segment's travel time TT_s is computed using Equation 2.

$$TT_s = \frac{1}{n}\sum_{i=1}^{n} TT^i \tag{2}$$

Data-driven TSE approaches for travel time estimation and prediction aim at leveraging the relationship (model) between the supply and the traffic demand at various road segments, to approximate any of the traffic variables.

2.2 Geospatial Indexing

Geospatial data depict geographical information such as longitudes and latitudes. Geospatial indexes are data structures developed for efficient handling, storage, retrieval and processing of data with spatial attributes, and they are developed from well-known structures such as sorted arrays, binary trees, B-trees, and hashing (Lu and Ooi, 1993).

One geospatial indexing approach is *geohash*. Geohash is a hierarchical spatial data structure, which subdivides spatial regions into bounding boxes or grid buckets at different granulation and precision levels (Niemeyer, 2019; Vukovic, 2016). Geohash uses a base-32 (32-bits) alphanumeric character encoding, to produce unique ASCII strings. This string serves as an identifier representing a bounding box containing specific GPS coordinates. Moreover, geohash is hierarchical hashing algorithm with twelve levels known as precision levels. Each level defines a bounding box size given a spatial region. The bigger the size of the bounding box, the larger the number of GPS points contained (La Valley et al., 2017). Table 1 presents details about the twelve geohash precision levels and their bounding boxes size.

Table 1: Geohash Precision Levels and Their Bounding Box Size (Levels One (1) to Seven (7)).

Geohash spatial indexing	
Precision level	**Bounding box area**
1	$\leq 5{,}000\ km \times 5{,}000\ km$
2	$\leq 1{,}250\ km \times 1{,}250\ km$
3	$\leq 156\ km \times 156\ km$
4	$\leq 39.1\ km \times 19.5\ km$
5	$\leq 4.89\ km \times 4.89\ km$
6	$\leq 1.22\ km \times 0.61\ km$
7	$\leq 153\ m \times 153\ m$

It is easy to move between the levels of precision. Higher precision levels have longer geohash codes and the bounding boxes containing them have a geohash string with the same prefix. For instance, the coordinates $x = (46.9466, 7.4426)$ are mapped to the level six geohash *um*716 and to the level seven geohash *um*7167. Figure 1 depicts the aforementioned coordinates at the two levels. Note the difference in the size of the bounding boxes, the smaller the bounding box, the higher the precision and longer the length of the geohash code.

(a) Level Six (6) Geohash for the Point x.

(b) Level Seven (7) Geohash for the Point x.

Figure 1: The Coordinates $x = (46.9466, 7.4426)$ Mapped to Geohash Codes in Levels Six (6) and Seven (7).

2.3 Related Works

Previous research efforts addressing the task of estimating and predicting travel time using probe data from logistic vehicles are presented in this section.

Zhang et al. (2017) constructed multiday spatiotemporal speed diagrams with probe data collected from logistic vehicles in Beijing, China. They made use of correlation traffic features in space and time by constructing a gray-level co-occurrence matrix (GLCM). A similarity measure was calculated with normalized square differences (NSD) between current and historical GLSMs, to select candidate traffic patterns. The future travel times were estimated by comparing current conditions to similar experienced travel times.

Another related initiative is one of Yoon et al. (2007). A novel approach for aggregating data temporally for an identified spatial region was introduced. Relying on probe data from a single taxi on a fixed road segment, their approach attempted to characterize traffic patterns and identify traffic states, as well to address the low sampling rate problem due to the limited number of vehicles per road segment. The authors found that the traffic patterns obtained by studying the behavior of the taxi were consistent over time, thanks to the segment-oriented analysis that they performed. Even though estimating and forecasting travel time was not the intention of this research work, it contributes to finding a solution to the low penetration/sampling rate problem of delivery vehicles.

In the work of Wang et al. (2014), the development of a real-time model for estimating travel time within a city was proposed. The researchers addressed the problems of data sparsity that working with probe data brings and responding quickly to users, by modeling travel times in different road segments and making use of three-dimensional tensors.

In contrast to the aforementioned research efforts, this work proposes a scalable approach to estimate and calculate near future travel times by using logistic probe data, pattern searching and geospatial indexing on temporally aggregated data, while having acceptable levels of accuracy.

3 METHODOLOGY AND USE CASE

The guidelines of the design science research for information systems methodology were followed in the development of this work. This research methodology was selected because its application entails the development of an artifact while extending existing knowledge (Hevner and Chatterjee, 2010). Moreover, the fact that this project was developed in collaboration with *Swiss Post*, the national postal company of Switzerland, eased the selection of the research methodology, as this company supports investigation but is also interested in obtaining practical solutions in the process.

The method in this study encompassed four main phases: i) data cleaning and selection; ii) travel time approximation; iii) travel time prediction; and, iv) evaluation. Figure 2 depicts theses phases as well as intermediate steps.

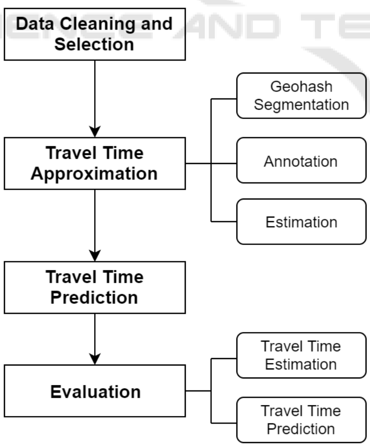


Figure 2: Methodology.

3.1 Data Cleaning and Selection

Data of five months of operations of logistic trucks of Swiss Post was used. The initial database consisted of 353.1 million records and each record was described in terms of twenty-six fields. The information described in the records corresponded to GPS location of the vehicles during their delivery routes, speed, mileage, events (e.g., parked, motor on and off), driver and vehicle identification, street name, among others. The operations took place from July to November of 2018 and only the data points registered on the area of Bern-Ostermundingen (Switzerland) were considered, given the limitations of our computational resources and as it was a familiar area for the *Swiss Post* representatives supporting this project. Moreover, duplicates, inconsistent and invalid records were removed.

3.2 Travel Time Approximation

The goal of this phase was to calculate travel time through the sum of cruising time within segments that constitute a journey.

The analysis carried out in this project was based solely on probe data from delivery vehicles. The data contained detailed timestamped location records, which are useful for deducing travel times between any two arbitrary points. However, interruptions caused by business activities (i.e., delivering a package) need to be properly handled as well as the low penetration rate of delivery vehicles in estimating travel times.

Regarding the usage of historical traffic data, two assumptions about road traffic and travel time were made:

1. Historical data contains latent traffic relationship valid for current and future traffic conditions. This assumption follows the general approach of pattern matching methods, which establishes that traffic patterns are recurrent in nature and therefore similar historical events can be used to provide estimates on current conditions (Zhang et al., 2017).
2. With a large amount of data from any given segment, an expected value for the travel time can be approximated with the average obtained from past trips in that segment. This assumption follows the strong law of large numbers (See Equation 3) (Loéve, 1997) .

$$Pr\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1 \tag{3}$$

which asserts that the probability Pr that the average of the observations converges to the expected value as the number of points n becomes larger, is equal to one.

An estimation model was derived based on the three phases of TSE methods (refer to Sec. 2.1).

3.2.1 Geohash Segmentation

The low penetration rate of probe vehicles per delivery routes (specifically 1 per route) in the area of Bern-Ostermundingen, led to finding ways to overcome this limitation; thus, temporal aggregation and spatial segmentation were applied. The temporal aggregation implied studying the behavior of the vehicle circulating on different days, whereas the spatial segmentation entailed dividing the space into segments and analyze the traffic data on a segment-to-segment basis, following common practices of TSE methods.

Geohash was the approach employed to segment the space, as it allows grouping points to a common spatial bin (or bounding box) of a fixed size. The chosen geohash level was eight meaning that the bounding boxes for the segmentation had a size of $\sim 38m \times 19m$, which was decided given the distribution of the GPS points and the level of detail that the researchers hoped to achieve.

3.2.2 Annotation

The annotation consisted of approximating the expected speed for each segment. This speed was calculated as follows:

For a segment r of length l (length deducted from the size of the geohashes), there is a travel time expectation TT_r that can be calculated in terms of the average time of past trips in the segment r. Thus, the expected (mean) speed $\bar{s}$ is calculated as expressed in Equation 4.

$$\bar{s} = \frac{l}{TT_r} \tag{4}$$

The speed $\bar{s}$ corresponds to the segment's mean speed; however, applying the strong law of larger number this mean speed is assumed to be the expected speed for the segment r.

3.2.3 Estimation

The expected travel time for a given segment was computed using the average travel time of past trips. To compute the actual travel time for a current trip, the real-time average speed in the segment s needs to be used. Thus, considering the length l of the segment, for a vehicle i, its actual travel time TT_i can be determined using Equation 5.

$$TT_i = \frac{l}{s} \tag{5}$$

Moreover, unusual traffic conditions produce deviations from the expected mean speed. Lower average speed than the expected segment's speed signals unfavorable conditions, while higher average speed suggests better traffic dynamics than usual. Thus, differences in the actual travel time and the expected travel time could be the result of changes in the traffic situation, stops due to pedestrian crossing, and driver behavior. This time difference is expressed in terms of ε (segment delay) and therefore, the actual travel time can be reformulated with Equation 6.

$$TT_i = TT_r + \varepsilon \tag{6}$$

The segment delay ε can be expressed using Equation 7.

$$\begin{aligned} \varepsilon &= TT_i - TT_r \\ \varepsilon &= \frac{l}{s} - \frac{l}{\bar{s}} \\ \varepsilon &= \frac{\left(s - \bar{s}\right) \times l}{s \times \bar{s}} \end{aligned} \tag{7}$$

Since l, TT_r and $\bar{s}$ are known values, the actual travel time is solely dependent on the current speed s which captures other road conditions. At each point in time, the instantaneous speed or the average of the recorded speed are used to calculate ε. It should be pointed out that negative error terms may occur as a consequence of possible favorable traffic conditions, and therefore, decreased travel times than the expectation.

3.3 Travel Time Prediction

Predicting arrival time at delivery targets epitomize travel time prediction in the delivery business. In terms of existing prediction models, there are four groups (Mori et al., 2015): i) *naive*, methods that do not model traffic data but make diverse assumptions to deliver a fast prediction; ii) *traffic flow-based*, techniques that rely on mathematical relations between traffic flow, density, and speed; iii) *data-based*, approaches that rely on historical data to find relationships between traffic variables; and, iv) *hybrid*, methods that combine concepts from the aforementioned groups.

The complexities in predicting travel with probe data are further aggravated due to the irregular and complex business activities, which involve numerous external waiting times besides the usual traffic behavior. As no particular model was found suitable in our case study, a hybrid approach was adopted and the following steps were followed.

For pre-processing purposes, a non-parametric pattern searching approach was used to filter out external waiting times. Estimations of segment delay (ε), expected travel time ($E[TT_r]$) and current travel time TT_i, were deduced using the expressions deducted and presented in Section 3.2.3.

To predict the travel time from a current segment r to a target segment t, two scenarios were considered: the first scenario (see Fig. 3), illustrates the case where multiple vehicles are present in segments of the same route. With live data, ε terms are calculated and a dynamic travel time prediction can be computed using Equation 8.

$$TT(r \rightarrow t) = \sum_{i=1}^{k} (E[TT_r]_i + \varepsilon_i) \tag{8}$$

where k is the number of segments within a prediction horizon from the current segment to the target. In simple words, the individual travel times and delays per each vehicle and segment are computed and summed out to predict the travel time to the target t.

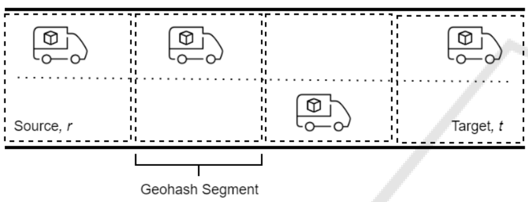


Figure 3: Prediction Scenario with Real-Time Information from Constituent Segments.

The second scenario (see Figure 4), corresponds to the case where only one vehicle is present in a particular route. A multi-step gradual approach is undertaken to compensate for the unknown future delays within segments. The prediction can be framed in terms of the sums of the delays ε within segments from a source to destination, given each segment's $E[TT_r]$.

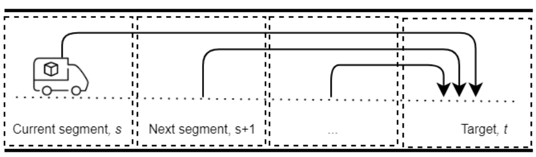


Figure 4: Multi-Step Gradual Prediction Approach. Delays from the Current Segment Are Used to Adjust Expected Travel Time to the Destination Segment.

At each current segment, real-time data is used to calculate the current ε term and the travel time is adjusted accordingly. ε terms for segments without information is assumed to be zero, and therefore the latest expectation values are used. Under these assumptions, the travel time is predicted using Equation 9.

$$TT(s \rightarrow t) = \varepsilon_s + \sum_{j=1}^{k} E[TT_r]_j \tag{9}$$

where s indicates the current segment and its corresponding delay term ε_s, j represents subsequent segments from the current segment until the destination segment. This approach, however, takes into consideration only cruising time. Additional waiting times in each segment are neglected, thus, the closer the target is approached, the more accurate the prediction.

3.4 Evaluation

The evaluation stage consisted of assessing the reliability of the deduction of the segment speed expectation and the travel time estimation and prediction. For the evaluation, the dataset was split into two parts:

- **Historical Data.** This is the dataset used for the artifact prototyping and model refinements. The historical data consisted of data recorded during a four-month delivery period, from July to October. The 4-month data contained 85% of the total cleaned data (approx. 140,000 records).
- **Test Data.** A separate dataset was prepared for testing purposes only and not used during the modeling process. It consisted of delivery data for the month of November, being approximately 15% of the total cleaned dataset.

3.4.1 Travel Time Estimation

Given that records in the dataset contained a field with instantaneous speed, to provide evaluations for the deduced segment expected speed, the mean average percentage error measure (MAPE) was chosen to compare the accuracy of the travel time forecasting using the instantaneous speed and the deduced segment mean speed from our approach. MAPE is a dimensionless measure and a common approach for comparing different forecasting models (Zhang et al., 2017). MAPE expresses the magnitude of the error relative to the ground truth as a percentage and is defined using Equation 10.

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \times 100\% \tag{10}$$

where n is the number of observations, A_t is the actual value and f_t is the forecasted value. Lewis (1982) defined four ranges with their interpretations for typical MAPE values found in industrial and business data: an error smaller than 10%, then the forecasting is accurate; values between 10% and 20% indicate that the forecasting is good; values between 20% and 50% show that the forecasting is not inaccurate but not good either, and finally, a value greater than 50% indicates that the forecasting is inaccurate.

3.4.2 Travel Time Prediction

To asses the accuracy of the proposed prediction model, a naive model was implemented using the test dataset and served as a baseline for comparison. Furthermore, the metrics of MAPE and the mean absolute error metric (MAE) (see Equation 11) were used to measure the magnitude error in time from the prediction models.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \tag{11}$$

where A_t corresponds to the prediction and F_t to the true value.

4 RESULTS

4.1 Data Cleaning and Selection

Once the data cleaning took place, it was found that 45% of the records were duplicates or were not useful for the goals of this work, resulting in a dataset of 315,014 records. Moreover, after discarding fields that were not relevant for analysis, each record was described in terms of fourteen fields.

4.2 Travel Time Approximation

After the segmentation, annotation and estimation steps were executed, additional fields were added to the records in the dataset: i) *geohash*, the string referencing the coordinates; ii) *distance*, in kilometers until the next destination point in the next segment; iii) *duration*, the time employed to reach the next segment point; iv) *waiting time*, the time elapsed at a location within a segment; and v) *mean speed*, the deduced speed calculated applying the Equation 4.

Furthermore, the MAPE was calculated to compare the results of the travel time estimation using the instantaneous speeds (from the historical data) and the deducted expected mean speeds. Relying on the deduced mean speed, the MAPE value of 6.03 was obtained, suggesting that our approach is near highly accurate; relying on the instantaneous speed, the MAPE value of 128.68 is obtained which signify inaccurate estimates.

These results signal that instantaneous speed does not provide an accurate generalization for the mean speed within segments. As such, the estimation using point values results in overestimation or underestimation and consequently, in reduced accuracy in the travel time estimates.

4.3 Travel Time Prediction

The travel time prediction accuracy was evaluated applying the MAPE and MAE metrics. The proposed approach and the defined baseline were compared to the dataset values. As per the results presented in Table 2, this approach had a reasonable forecasting performance (MAPE value of 23.6), with a mean absolute error of fourteen minutes and thirty-three seconds. In contrast, the baseline (naive model) had a poorer performance.

Table 2: MAE and MAPE Travel Time Prediction Accuracy Comparison between the Baseline and the Proposed Approach.

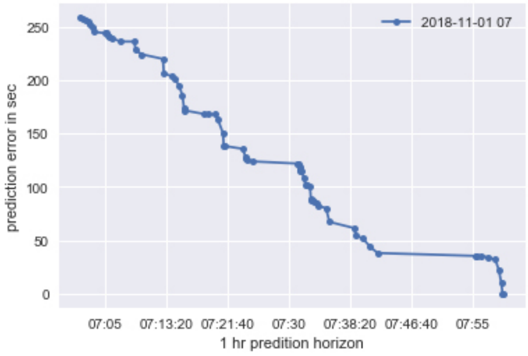
	MAE	**MAPE**
Baseline	39” 08’	43.9
This Approach	14” 33’	23.6

Although the MAE measurement of this proposed approach seems daunting, the authors argue that it is a very acceptable result considering the simple methods that were used. Moreover, when calculating the MAE values from segments closer to the destination, they decreased, meaning that the forecasting becomes more accurate as illustrated in Figure 5 in two randomly selected trips.

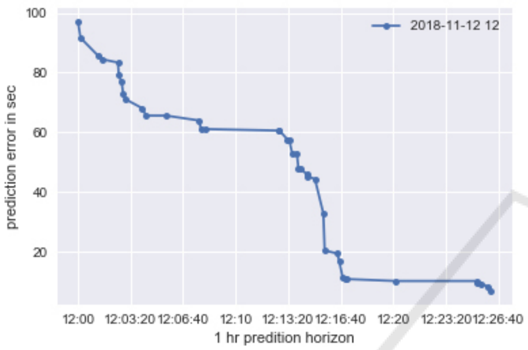
5 SUMMARY AND CONCLUSIONS

This research work proposes a data-driven approach to forecasting travel times of en-route logistics vehicles. Data on the operations of a *Swiss Post* delivery vehicle were analyzed. By means of the design science research methodology, an artifact was implemented following a procedure that consisted of four stages: i) *data cleaning and selection*, ii) *travel time approximation*, iii) *travel time prediction*, and iv) e*valuation*.

The data cleaning and selection stage allowed the removal of inconsistent records and disregard fields that were not of interest. After this stage was completed, the dataset was composed of 315,014 records which were studied. Later, the travel time approximation stage took place, which entailed aggregating the data of the different days and the intermediate steps of segmentation, annotation, and estimation. The segmentation step was conducted through the application of geohash to perform a segment-to-segment analysis of the behavior over time of the logistic vehicles; following, the annotation step encompassed estimating the expected mean step for each segment; lastly,



(a) Date: 2018-11-01 from 07:00 to 08:00.



(b) Date: 2018-11-12 from 12:00 to 13:00.

Figure 5: Prediction Error Patterns in an Hour Interval Denoting Source and Destination Segments from Two Randomly Selected Trips on Different Days and Different Periods.

the estimation step implied deducting the expressions that allowed estimating the travel time of a vehicle for a current trip. Afterward, the travel time prediction stage took place, whose purpose was to define the expressions that allowed forecasting a vehicle's arrival travel time to a destination. Finally, the evaluation stage aimed at assessing the accuracy of the estimation and prediction phases by calculating MAPE and MAE metrics.

The results of the evaluation stage suggest that this approach is a feasible implementation, as they were nearly accurate and showed higher accuracy than the baseline methods. Even though the MAE value showed that there was a difference of fourteen minutes in average between the prediction and the historical data, the authors consider that these results are satisfactory considering the low penetration rate of the probe vehicles studied.

Moreover, this initiative differentiates itself from other methods that rely on map-matching as geohashing requires low computational power in comparison to the computationally intensive map-matching algorithms. In addition, the geohashing indexing allows concurrent modeling and analysis at different levels, in a course-to-fine manner and vice-versa, which facilitates analysis tasks.

In terms of the travel time expectation and prediction computation, this approach employs rather simplistic models that are low in complexity and easily scalable unlike other methods based on machine learning algorithms which require massive resources. The authors argue that efforts as this one are promising alternatives that deserve to be explored towards developing less complex and more efficient solutions.

Furthermore, the results obtained in this work could be translated to improvement in the quality of delivery services and even the development of new ones. Besides, segment-to-segment and granular analyses allow getting detailed insights about what happens on the roads, that can serve as a basis to optimize the routing of vehicles and therefore, fewer resources consumption.

To close the curtains on this research effort, it should be highlighted that the expectation of traffic conditions is time and context-dependent. For example, during rush hours one expects less favorable traffic conditions and therefore, the expectation for trips at rush hour and non-rush hour periods should be modeled differently. Furthermore, weather conditions and special events (e.g., concerts and public demonstrations) incise the time needed to reach a destination. Future improvements to this initiative will be directed towards incorporating such contextual information, to improve the developed models and provide more accurate results.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the Human-IST Institute at the University of Fribourg for contributing with valuable thoughts and comments. We especially thank the Secretariat of High Education, Science, Technology, and Innovation (SENESCYT) of Ecuador and the *Swiss Post* for their support to conduct this research.

REFERENCES

D'Andrea, E. and Marcelloni, F. (2017). Detection of traffic congestion and incidents from gps trace analysis. *Expert Systems with Applications*, 73:43–56.

Hevner, A. and Chatterjee, S. (2010). *Design research in information systems: theory and practice*, volume 22. Springer Science & Business Media.

La Valley, R., Usher, A., and Cook, A. (2017). Detection of behavior patterns of interest using big data which have spatial and temporal attributes. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4.

Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth-Heinemann.

Lin, H.-E., Zito, R., Taylor, M., et al. (2005). A review of travel-time prediction in transport and logistics. In *Proceedings of the Eastern Asia Society for transportation studies*, volume 5, pages 1433–1448.

Loéve, M. (1997). *Probability Theory, Vol. I.* Springer, fourth edition edition.

Lu, H. and Ooi, B. C. (1993). Spatial indexing: Past and future. *IEEE Data Eng. Bull.*, 16(3).

Mori, U., Mendiburu, A., Álvarez, M., and Lozano, J. A. (2015). A review of travel time estimation and forecasting for advanced traveller information systems. *Transportmetrica A: Transport Science*, 11(2):119–157.

Nanthawichit, C., Nakatsuji, T., and Suzuki, H. (2003). Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. *Transportation research record*, 1855(1):49–59.

Niemeyer, G. (2019). geohash.org. *Retrieved July*, 15.

Pu, W., Lin, J., and Long, L. (2009). Real-time estimation of urban street segment travel time using buses as speed probes. *Transportation Research Record*, 2129(1):81–89.

Ruppe, S., Junghans, M., Haberjahn, M., and Troppenz, C. (2012). Augmenting the floating car data approach by dynamic indirect traffic detection. *Procedia-Social and Behavioral Sciences*, 48:1525–1534.

Seo, T., Bayen, A. M., Kusakabe, T., and Asakura, Y. (2017). Traffic state estimation on highway: A comprehensive survey. *Annual Reviews in Control*, 43:128–151.

Treiber, M. and Kesting, A. (2013). Trajectory and floating-car data. In *Traffic Flow Dynamics*, pages 7–12. Springer.

Van Lint, J. and Van Hinsbergen, C. (2012). Short-term traffic and travel time prediction models. *Artificial Intelligence Applications to Critical Transportation Issues*, 22(1):22–41.

Vukovic, T. (2016). Hilbert-geohash-hashing geographical point data using the hilbert space-filling curve. Master's thesis, NTNU.

Wang, Y., Zheng, Y., and Xue, Y. (2014). Travel time estimation of a path using sparse trajectories. pages 25–34.

Wang, Z., Lu, M., Yuan, X., Zhang, J., and Van De Wetering, H. (2013). Visual traffic jam analysis based on trajectory data. *IEEE transactions on visualization and computer graphics*, 19(12):2159–2168.

Yoon, J., Noble, B., and Liu, M. (2007). Surface street traffic estimation. pages 220–232.

Yuan, J., Zheng, Y., Zhang, C., Xie, W., Xie, X., Sun, G., and Huang, Y. (2010). T-drive: driving directions based on taxi trajectories.

Zhang, Z., Wang, Y., Chen, P., He, Z., and Yu, G. (2017). Probe data-driven travel time forecasting for urban expressways by matching similar spatiotemporal traffic patterns. *Transportation Research Part C: Emerging Technologies*, 85:476–493.

Zhou, P., Zheng, Y., and Li, M. (2012). How long to wait?: predicting bus arrival time with mobile phone based participatory sensing.