Machine Learning in Vehicle Travel Time Estimation: A Brief Technological Perspective and Review

Son Pham
Universität der Bundeswehr
München, Germany
son.pham@unibw.de

Marian Sorin Nistor
Universität der Bundeswehr
München, Germany
sorin.nistor@unibw.de

Loi Cao
Le Quy Don Technical University,
Vietnam
loi.cao@lqdtu.edu.vn

Gerschberger Markus
University of Applied Sciences
Upper Austria
markus.gerschberger@fh-steyr.at

Maximilian Moll
Universität der Bundeswehr
München, Germany
maximilian.moll@unibw.de

Milani Rudy
Universität der Bundeswehr
München, Germany
rudy.milani@unibw.de

Abstract

A precise Estimated Time of Arrival (ETA) finds applications in various domains, such as navigation and logistics systems. This problem has gained a lot of attention from the research community. Machine learning has recently been applied and has shown promising results for ETA. Machine learning approaches can be divided into two categories, which are route-based and origin-destination-based methods. The first one divides the route into segments and predicts the ETA based on the information of these segments. The last one predicts ETA based on a few natural information, such as the origin, the estimation, and the departure time. In this paper, we aim to review recent studies of the mentioned machine learning approaches for ETA to determine the necessary input for an ETA forecasting model, the critical factors, and suitable approaches for ETA. Furthermore, we will discuss promising research directions to improve ETA, such as formulating ETA as a time series forecasting problem, including uncertainty or using ensemble learning models.

Keywords: ETA, Machine Learning, Route-based ETA, Origin-Destination-based ETA

1. Introduction

"One of the most critical location-based services (LBS) is the Estimated Time of Arrival (ETA) or vehicle Travel Time Estimation (TTE) (Z. Wang et al., 2018)". As a crucial component of many systems, such as navigation and intelligent transportation systems, it is becoming increasingly important and prevalent (Li et al., 2018). Applications for ETA can be found in ride-hailing, logistics, and shipping, where the duration of the trip substantially impacts the service's quality. A precise ETA will also improve the transportation system's efficiency in reducing negative externalities such as user travel costs, energy usage, and motor vehicle pollution. ETA has consequently created a crucial element that affects decision-makers (Z. Wang et al., 2018).

Route-based and origin-destination-based methods are the two main approaches for estimating the arrival time for road transport (Li et al., 2018; Z. Wang et al., 2018). There are publications on ETA prediction in maritime and air transport but this paper focuses only on road transport. The selected publication are some of the most recent and impactful one.

On the one hand, route-based techniques need route knowledge to develop a forecast. A route can be considered as a collection of successive segments. The total travel time on a route is the summation of the duration spent on each segment.

On the other hand, origin-destination-based approaches can forecast travel time even without route data. For forecasting ETA, just a few numbers of raw input features are provided, including the origin, destination, and departure time (Li et al., 2018).

In this report, we aim to review recent studies of the above approaches to clarify the following research questions:

- What kinds of data are needed as the input for the ETA forecasting model?
- What might impact an ETA?
- What are suitable approaches for ETA?

The paper is structured as follows. Section 2 discusses the ETA forecasting approaches based on machine learning. Section 3 outlines some promising future research directions, and Section 4 summarizes our work.

URI: https://hdl.handle.net/10125/106552
978-0-9981331-7-1

2. Machine Learning Approaches for ETA

This section will discuss two main categories of machine learning-based ETA approaches: route-based and origin-destination-based. For clarity, we divided the route-based approaches into two sub-categories: the route-based stand-alone and the hybrid approach. Figure 1 illustrates the taxonomy of machine learning approaches for ETA forecasting.

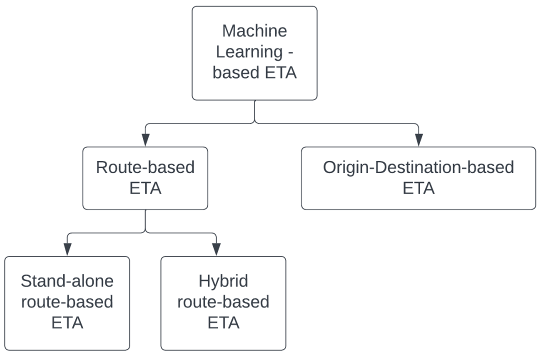

Figure 1. Taxonomy of machine learning approaches for ETA

2.1. Stand-alone Route-based Approach

In route-based ETA methods, the methods first estimate travel time for each segment on the individual link and then sum up all travel time to create a total travel time of the trip. It often accumulates local errors, and external features are often ignored. Researchers such as Z. Wang et al. (2018), Xu et al. (2022), and Yuan and Li (2019) have done many works on the ETA topic to address these drawbacks.

Z. Wang et al. (2018) had established a comprehensive, efficacious feature system for the local-based data. They gathered information for the Beijing ride-hailing service. The study conceptualized ETA as a classical regression problem that can be solved using deep learning (DL) and traditional machine learning (ML) techniques. The authors proposed the Wide-Deep-Recurrent model and compared them with some of the most well-known ones from the literature, such as TEMP (H. Wang et al., 2019) or PTTE (Y. Wang et al., 2014).

From the literature, traveling data are categorized into five followings groups:

- Spatial Information: Spatial information is critical because travel time correlates positively with the route. Spatial data includes the features of all the building blocks that form the route, such as the road segment, traffic congestion, and traffic light information. For instance, the features can be represented by the segment's length, width, quality, and grade. Furthermore, some other exciting features can be the number of lanes, the index number of the segment, as well as the points of interest (POIs) information on the route (Z. Wang et al., 2018).

- Time-related Information: Time-related data is another crucial factor influencing vehicle travel time during rush or off-peak hours. Thus, the trip's departure time can be represented with different features, including the season or special events such as holidays or summer vacation, and rush hours indicator (Z. Wang et al., 2018). This information can generally be represented as a time series.

- Traffic Information: The traffic network's circumstances directly impact travel time. Traffic data can be defined as the travel speeds for each road segment. There are different speeds: average, real-time estimated, and free-flow speed (Z. Wang et al., 2018).

- Behavior Information: Every driver's travel time may differ. Thus, personalized features include driver, rider, and vehicle profiles (Z. Wang et al., 2018).

- Extended Information: Other open data is used as improved features and enhances the precision of the ETA forecast. They can be weather forecast and traffic restrictions (Z. Wang et al., 2018).

Figure 2 shows the overview of all critical route information.

The data was collected in Beijing on the DiDi platform from January to May 2017. There are two kinds of data regarding the operational status of the driver, namely pickup and trip information. On the one hand, a pickup record is accumulated when the driver reacts to a rider's request until they pick up the rider. On the other hand, the time when the driver drives the passenger to his destination represents a trip sample (Z. Wang et al., 2018). After removing the abnormal cases, the data contained about 57 million pickup and 62 million trip samples. However, the data is not publicly available because of privacy concerns. Therefore, building comprehensive "features for the location-based data" and establishing a "high dimensional feature mapping" for them is one of the main contributions of (Z. Wang et al., 2018).

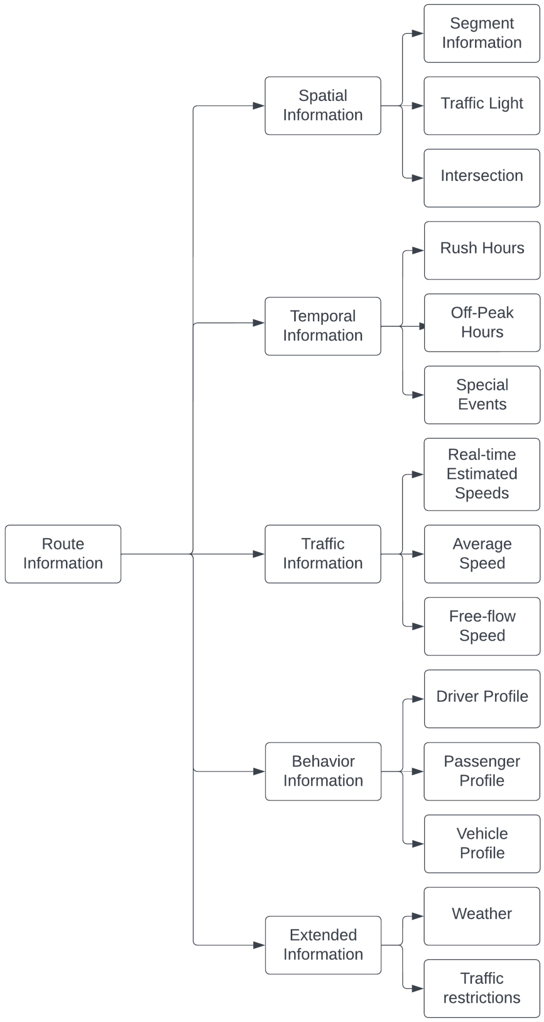

Figure 2. Overview of route information

The authors then defined ETA as a regression problem on the extracted data. Both traditional ML (the gradient boosting decision tree - GBDT and factorization machines – FM) and their proposed DL method (Wide-Deep-Recurrent - WDR), "a jointly trained model between wide linear models, deep and recurrent neural networks together" (Z. Wang et al., 2018), were applied to solve this problem. The experimental results demonstrated that their proposed model, WDR, outperforms traditional techniques, GBDT and FM, and two state-of-the-art methods – Temporally weighted neighbors (TEMP) (H. Wang et al., 2019) and Path Travel Time Estimation (PTTE) (Y. Wang et al., 2014) on both the pickup and trip-based data. Temporal speed reference is used by TEMP to assign weights to adjacent trips (H. Wang et al., 2019) while PTTE is based on spare trajectories produced by a sample of vehicles both in the past and in the present (Y. Wang et al., 2014). Furthermore, the travel time may vary in the future because the exact route is unknown before the trip starts. In order to compare the performance of the two models, WDR-GPS (trained on historical GPS trajectories) and WDR-RP (trained on planning routes), the authors used historical GPS trajectories as their training data. Results from testing a dataset with characteristics extracted along the planned routes revealed that WDR-RP outperforms WDR-GPS significantly.

Xu et al. (2022) attempts that either disregarded or assumed that the mode information known for each training and querying trajectory would be relevant to the issue of the impact of transportation modes on travel time estimation (TTE). The authors proposed a multi-task learning model (MTLM) for solving a TTE problem. MTLM consists of two tasks: a supervised learning task to recommend transportation mode and an unsupervised learning task for preserving related data and TTE of a given path. Additionally, the authors combined the pertinent data connected to the suggested modes of transportation, such as trajectory data and other outside variables (holidays, weekdays, or weekends). This information is utilized to estimate travel duration together with spatial correlations, temporal dependencies, and the influence of the method of transportation. MTLM was tested using the GeoLife dataset that Microsoft Research published (Zheng et al., 2008; Zheng et al., 2010; Zheng et al., 2009). The experimental results confirmed that LTLM performed well and often better than well-known models, such as the one described below.

2.2. Hybrid Route-based Approach

Besides the above approaches, the work of Hu et al. (2022) and D. Wang et al. (2018) attempted to establish a hybrid approach for ETA. The hybrid approach uses different models for each step of the estimation process instead of a single model for the whole process. The goal of Wang's study (D. Wang et al., 2018) was to solve the shortcomings of the route-based strategy, such as failing to account for traffic signals, road intersections, and potentially accumulating local errors. The authors proposed an end-to-end deep learning framework for the whole path's Travel Time Estimation (DeepTTE). The DeepTTE is trained on spatial/temporal data and historical trajectories for predicting future ETA with only path data. In order to capture spatial correlations, they describe a geo-convolution procedure incorporating geographic data into the traditional convolution. During the training phase, the proposed model acquires the ability to simultaneously estimate the travel times of each segment and the full path. On the Chengdu dataset and the Beijing dataset, the proposed model was put to the test. The authors incorporated spatial, temporal, and environmental (traffic lights, weather) elements into the model in this work. The experiments show encouraging outcomes.

Hsu, 2021 presented the HyperETA ML-based ETA method to estimate ETA for a given trajectory. HyperETA used a hypercube-clustering technique to assess the similarity of the data from multiple trajectories. The main idea behind hypercube clustering is to depict trajectories using hypercubes that are more resistant to noisy trajectories. From previous trajectories, HyperETA creates a trajectory model. The travel time for a specific trajectory (converted into hyper-cubes) is then estimated using the trajectory model. After adding up these times, the trip's ETA is calculated. The authors used Chengdu taxi trajectory data as a benchmark for research comparing HyperETA and DeepTTE. The outcomes showed that, in terms of prediction accuracy, HyperETA outperformed DeepTTE.

Recently, Hu et al. (2022) introduced a hybrid approach between a deep residual ETA network (DeeprETANet) and a routing engine ETA model. In this hybrid, ETA's routing engine will learn to predict ETA with features extracted from real-time traffic and maps. While DeeprETANet learns to reduce the discrepancy between the ETA of the routing engine and the actual arrival time. The proposed model was evaluated on a global request dataset (ride-hailing and eats delivery) from Uber's platform in 2021. The dataset is constructed from context, temporal, spatial, trip, and traffic features. The experimental results demonstrate that DeeprETANet outperforms the routing engine ETA and baseline regression models. Instead of predicting ETA from location-based data, the work in Yuan and Li, 2019 focused on the trajectory data analytic task to benefit ETA prediction methods. The authors first design a road-network-aware trajectory similarity function to quantify trajectory similarity. Then, to use the function for trajectory search and join operations, they proposed a filtering-refine framework in which the filtering step will prune many dissimilar pairs of trajectory points. In contrast, the refining step will verify the trajectory candidates after pruning.

2.3. Origin-destination-based Approach

Origin-destination-based ETA methods (OD-ETA) aim to estimate the travel time without the actual route information (Li et al., 2018). This approach is suitable for online estimation services where only the origin and the destination are given before a trip (Jindal et al., 2017; H. Wang et al., 2019). Note that the road network information is not available during the testing phase. However, origin-destination-based methods have suffered from several challenges. First, without the path information, the raw input attributes used for online prediction are limited to the origin, destination, and departure time. Second, using natural traits as part of a prediction model does not seem easy. It is difficult to estimate the distance or degree of resemblance between two trips using simply the latitudes and longitudes of their starting and ending points. We will discuss recent research of Li et al. (2018), H. Wang et al. (2019), and Yuan et al. (2020) for this approach.

A paramilitary study on the OD-ETA approach can be known as the study of H. Wang et al. (2019). In this work, H. Wang et al. (2019) proposed a method based on nearest neighbors that calculates the travel time of the trip by averaging the scaled travel times of all previous journeys with comparable origins and destinations.

On the other hand, Li et al. (2018) attempted to learn better representation from limited raw features (the origin, the destination, and the departure time). The authors introduced a novel representation model to learn road networks and spatiotemporal information from historical data. Similar representations should be used for timestamps and locations close to each other. The authors extract diverse summaries from the path, such as the travel length, the number of traffic lights, the number of segments in the journey, and the number of turns, and operate them as additional tasks to be predicted rather than using the route information as extra input features that are inaccessible during testing (Li et al., 2018). A multi-task learning framework (MURAT) is specifically suggested to learn the primary job, forecasting ETA, and multiple auxiliary tasks, predicting other path summaries together. The authors employed two well-known datasets, BJS-Pickup and NYC-Trip, to evaluate MURAT. The source, destination, departure time, journey time, and path are all included in the BJS-Pickup dataset. The extra tasks extract various path summaries, including trip time, the quantity of roads, lights, and turns (Li et al., 2018). OpenStreetMap and Mapmatching are used for constructing a link graph. The results show that MURAT performs very efficiently on both datasets and outperforms other well-known machine learning methods. Similarly, the study Yuan et al. (2020) also constructs a latent representation learning the correlation between natural features (origin, destination, departure) with representing trajectories from historical trips.

3. Promising ETA Research Directions

One of the most promising directions to improve the prediction accuracy of ETA is to include time series forecasting models. The recent approaches consider ETA as a classical regression problem. However, ETA can also be defined as a time series forecasting problem by modeling the remaining time as a time-dependent value. Two of the most successful models for time series forecasting are Long Short Term Memory RNN (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Unit RNN (Chung et al., 2014).

The second idea to achieve better ETA is to include uncertainty in the prediction models. If the traditional machine learning pipeline is implemented, where the trained models are deployed to predict the ETA, uncertain factors such as traffic conditions, weather, and carrier performance are rarely considered. Measuring and including these factors in the prediction pipeline may be practical as a confidence interval or probability distribution. The real-time data on these factors can be gathered using the APIs of popular service providers.

The next conceivable direction of ETA research is to use ensemble learning. Ensemble learning in forecasting combines diverse predictive models to reach higher performance. Much research has proved that the ensemble model performs better than any single model (Breiman, 1996a; Freund et al., 1999). For ETA prediction, three major ensemble learning algorithms can be used, namely bagging (Breiman, 1996a), boosting (Freund et al., 1999), and stacking (Breiman, 1996b). Interested readers are referred to the cited references. Some authors already had the idea of averaging the ETA of different models to conduct the final ETA. It is the most straightforward form of ensemble learning. It is worth mentioning that besides the self-developed forecasting models, existing APIs such as the Distance Matrix API from Google can also be used as the input for ensemble models.

Last but not least, a data sharing system for traveling data could boost the motivation of the research community to develop new solution approaches. Traveling data may contain a lot of personal data. That is also why few good data sets are available for researchers.

4. Conclusion

Based on the discussion above, ML and DL can be powerful methods for estimating the arrival time in route- and origin-destination-based approaches. In route-based approaches, stand-alone and hybrid ML methods are utilized for ETA, in which ETA problems can be considered regression/clustering/classification problems. Such methods will predict ETA for a given path.

On the other hand, OD-ETA approaches can predict travel time for a given trip with natural features (weather information and traffic restrictions). OD-ETA often addresses the ETA problem by learning multi-tasks simultaneously, in which the main task learns to predict ETA while auxiliary tasks attempt to extract trajectories and other factors from historical data. Both approaches have their limitations. Route-based ETA methods often suffer because a querying trajectory in the future (not known at the departure time) can be different from the planning route. Thus, a trained model on historical trajectories may perform poorly on future trips with different trajectories from planners. On OD-ETA methods, the main difficulty is that only limited features are available for constructing prediction models.

Each approach works on its assumption; thus, training and querying data for each approach are quite different. For example, in the first approach, the training data and querying data often include features such as:

- Spatial information.
- Historical data: traffic information (speeds on the segment), trajectory information (lat, long, timestamp), and other features extracted from online services (OpenStreetMap).
- Other factors: temporal information (days of week/month/year, special days, time of days rush hour); the weather information and traffic restriction; driver profile.
- Raw features: the origin, the destination, and the departure time.

The training data often includes all the above features, while testing features do not include historical data. For the OD-ETA methods, the training data could be similar to the route-based methods. However, the testing phase requires only the natural features, and the path and trajectory information can be extracted from historical data using representation learning models.

References

Breiman, L. (1996a). Bagging predictors. *Machine learning*, *24*, 123–140.

Breiman, L. (1996b). Stacked regressions. *Machine learning*, *24*, 49–64.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, *14*(771-780), 1612.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hsu, O. L. (2021). Hypereta: An estimated time of arrival method based on hypercube clustering.

Hu, X., Binaykiya, T., Frank, E., & Cirit, O. (2022). Deepreta: An eta post-processing system at scale. *arXiv preprint arXiv:2206.02127*.

Jindal, I., Chen, X., Nokleby, M., Ye, J., et al. (2017). A unified neural network approach for estimating travel time and distance for a taxi trip. *arXiv preprint arXiv:1710.04350*.

Li, Y., Fu, K., Wang, Z., Shahabi, C., Ye, J., & Liu, Y. (2018). Multi-task representation learning for travel time estimation. *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 1695–1704.

Wang, D., Zhang, J., Cao, W., Li, J., & Zheng, Y. (2018). When will you arrive? estimating travel time based on deep neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1).

Wang, H., Tang, X., Kuo, Y.-H., Kifer, D., & Li, Z. (2019). A simple baseline for travel time estimation using large-scale trip data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(2), 1–22.

Wang, Y., Zheng, Y., & Xue, Y. (2014). Travel time estimation of a path using sparse trajectories. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 25–34.

Wang, Z., Fu, K., & Ye, J. (2018). Learning to estimate the travel time. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 858–866.

Xu, S., Zhang, R., Cheng, W., & Xu, J. (2022). Mtlm: A multi-task learning model for travel time estimation. *GeoInformatica*, 1–17.

Yuan, H., & Li, G. (2019). Distributed in-memory trajectory similarity search and join on road network. *2019 IEEE 35th international conference on data engineering (ICDE)*, 1262–1273.

Yuan, H., Li, G., Bao, Z., & Feng, L. (2020). Effective travel time estimation: When historical trajectories over road networks matter. *Proceedings of the 2020 acm sigmod international conference on management of data*, 2135–2149.

Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W.-Y. (2008). Understanding mobility based on gps data. *Proceedings of the 10th international conference on Ubiquitous computing*, 312–321.

Zheng, Y., Xie, X., Ma, W.-Y., et al. (2010). Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, *33*(2), 32–39.

Zheng, Y., Zhang, L., Xie, X., & Ma, W.-Y. (2009). Mining interesting locations and travel sequences from gps trajectories. *Proceedings of the 18th international conference on World wide web*, 791–800.