Data-driven Methods for Travel Time Estimation: A Survey

Zhipeng Zheng, Yongchao Ye, Yuanshao Zhu, Shiyao Zhang, *Member, IEEE*, and James J.Q. Yu, *Senior Member, IEEE*

***Abstract*— Travel time estimation is a crucial component of intelligent transportation systems, affecting various applications such as navigation, ride-hailing, and route planning. Traditional methods for travel time estimation rely on subjective judgments, limited data sources, and straightforward modeling techniques. Owing to recent advances in data mining and machine learning, numerous data-driven methods are adopted to address the problem that occurred in traditional schemes, which demonstrate exceptional performance. In this paper, we present a comprehensive survey of data-driven methods for travel time estimation, encompassing application scenarios, spatial-temporal modeling approaches, and data representation learning techniques. To support and promote further research in this field, we provide a valuable list of open data sources and source codes, offering researchers a solid foundation for their future endeavors. Furthermore, this survey discusses emerging trends and key challenges faced by the research community, such as the integration of real-time data streams and the use of uncertainty estimation. We also explore the potential impact of these advancements on transportation systems, highlighting opportunities for improvement and innovation. To the best of our knowledge, this work is among the first to offer a comprehensive, in-depth review of data-driven methods for travel time estimation, providing researchers and practitioners with a valuable reference in the field.**

I. Introduction

Travel time estimation (TTE) is a critical component of intelligent transportation systems (ITS), which plays a pivotal role in many location-based services, including navigation, ride-hailing, and route planning. Particularly, the proliferation of sharing economy, online ride-hailing, and car-sharing mobile apps has dramatically amplified the demand for accurate travel time estimates. This information allows individuals to create effective schedules, circumvent congested routes, and save time. Additionally, traffic management can leverage specific travel time data to evaluate road planning and enforce traffic control measures, thereby enhancing transportation efficiency.

Nevertheless, considering the complex spatio-temporal information and numerous factors that subtly influence travel time, TTE is a challenging task. A multitude of latent factors, such as meteorology, points of interest (POI), and personalized driving styles, can affect the accuracy of these estimates. Since the pioneering work in this field was published in 1973 [1], TTE has garnered considerable attention from the intelligent transport community, resulting in the development of an array of predictive methods from multiple perspectives over the years.

One straightforward solution is to utilize speed detectors, like loop sensors, to provide information on the travel speed of each road segment, thereby allowing for the calculation of travel time for each segment. However, the adoption rate of loop sensors in many cities is limited, and their large-scale installation is impractical due to high costs. In addition, the entire travel time cannot be simply derived by summing up the travel times of each individual road segment as it neglects the time to pass intersections.

Thanks to the abundant crowd-sourced trajectory data sourced from vehicles, numerous data-driven methods have been proposed over the past decades to address the challenge presented by static speed detectors. For instance, a method proposed in [2] relies on the weighted average travel time of neighboring trips, defined as those originating and ending in geographically similar locations. Despite its ability to provide quick estimations, this method struggles with predicting both longer and shorter trips. Longer trips suffer from a lack of sufficient neighboring trip data, while shorter trips are heavily influenced by dynamic traffic factors such as traffic lights. The absence of detailed route information further complicates this method, leading to significant errors, while its performance is also hindered by potential issues such as data sparsity [3].

The above heuristic-based solutions rely heavily on human knowledge to model the problem explicitly, not fully utilizing the wealth of information in the massive data. To address this challenge, an increasing number of end-to-end models have emerged, driven by advances in deep learning techniques. Empirical studies have confirmed the superiority of deep learning methods for their better performance in capturing latent features and modeling realistic traffic environments with dynamic changes. For instance, Wang *et al.* [4] propose a DeepTTE model, which independently captures spatial and temporal dependencies and outperforms contemporary baseline methods.

In recent years, the advancement of tensor decomposition algorithms and deep neural networks has significantly improved the accuracy of TTE. However, despite these improvements, TTE still faces numerous unresolved challenges. While there are existing reviews on bus travel time estimation [5] and deep learning methods for other noteworthy issues in ITS (e.g., traffic flow prediction [6]), to the best of our knowledge, this work is one of the first efforts to comprehensively review data-driven methods for travel time

This work is supported by the Stable Support Plan Program of Shenzhen Natural Science Fund (No. 20220815111111002). The authors are with the Research Institute for Trustworthy Autonomous Systems and Department of Computer Science and Engineering, Southern University of Science and Technology, China. James J.Q. Yu is with the Department of Computer Science, University of York, United Kingdom. Shiyao Zhang and James J.Q. Yu are the corresponding authors.

estimation for general vehicles. The primary aim of this work is to review TTE methods from multiple perspectives and bridge the existing gap in this field. We also provide our insights on current challenges and potential future research directions.

The rest of this paper is organized as follows. Section II introduces the formulation and basic definitions of TTE. Section III presents a categorization of the TTE problem based on the application scenario and route representation. Section IV provides a classification of TTE models based on the adopted data-driven techniques. Experimental data and resources are outlined in Section V. Section VI offers our insights on the current challenges and future directions.

II. Problem Formulation

Pang *et al.* [7] consider travel time estimation as a multi-step-ahead prediction problem for time series, whereas the prevailing perspective in current research regards it as a spatial-temporal regression problem. In deep learning-based methods, the goal of travel time estimation is to train a model that extracts influential features from the given data, and subsequently use this model to estimate the travel time, denoted as $\mathcal{T}$, for a given path, denoted as $\mathcal{P}$, during the test phase. This process can be formulated as follows:

$$\mathcal{T} = \mathcal{F}(\mathcal{X}, \mathcal{W}), \tag{1}$$

where $\mathcal{F}(\cdot)$ represents the travel time estimator, $\mathcal{X}$ denote the input of the estimator, and $\mathcal{W}$ represents the corresponding estimator parameter.

In real-world scenarios, travel time estimation service providers often only know the start and end points of the trip when they receive a query. Furthermore, privacy-preserving policies may restrict access to complete trajectory information, leaving only the origin and destination points available for analysis. For example, the largest public taxi data released by New York City [8] does not include intermediate GPS points. To overcome this challenge, researchers have developed methods for estimating travel time based solely on the origin-destination (OD) pair information, a process known as OD Travel Time Estimation. The goal of OD travel time estimation is to estimate travel time using the trip dataset, given the origin location, destination location, and departure time.

This work proposes several fundamental definitions related to the task of travel time estimation:

Definition 1 (Road Network): The definition of a road network can be broadly classified into two categories according to prevailing research on the issues mentioned above. A road network can be defined as a set of connected road segments, each represented as a directed edge [9]. When road intersections are treated as links, as in [10], it seems more reasonable. Alternatively, road networks can be defined using directed [11]–[13] or undirected graphs $G(\mathcal{V}, \mathcal{E})$ [14], where the vertex set $\mathcal{V}$ denotes all the road intersections and the edge set $\mathcal{E}$ refers to all road segments.

Definition 2 (Trajectory): A trajectory is a sequence of spatio-temporal sample points generated by the movement of vehicles, including location (longitude and latitude) and time information.

Definition 3 (Path/Route): A path or route is a sequence of connected road segments on the corresponding road network, leading from the origin to the destination.

Definition 4 (Trip): A trip is defined as a tuple $x_i = (o_i, d_i, t_i)$ with three key components: the origin location (o_i), the destination location (d_i), and the departure time (t_i). Additional optional components may represent external features that could influence the estimation.

III. Travel Time Estimation

A. Application Scenario

In this section, we divide the related works of this problem into two parts according to their application scenarios, which are car travel time estimation and bus travel time estimation.

1) Car Travel Time Estimation: Apart from the external factors considered at a macroscopic level (e.g., meteorology), the incorporation of personalized information (e.g., individual driving patterns) has shown promise in improving estimation accuracy for individualized (customized) travel time. Gao *et al.* [10] harnessed the inertial data collected from smartphones to identify aggressive driving events and fused them within a deep recurrent neural network for personalized travel time estimation. Reference [15] captured the similarity among drivers who share close driving preferences (i.e., possess similar driving styles) to transfer knowledge from the drivers with dense trajectory data to others with sparse data and, as a result, effectively alleviated the driver data sparsity problem.

Existing travel time estimation methods primarily rely on historical trajectory data, which may fail to make prompt responses to stochastic events, such as traffic accidents, road maintenance, and traffic controls. As an important source of accurate and reliable traffic information, urban surveillance systems can be leveraged to infer the real-time traffic state, which benefits a more precise travel time estimation. Building upon this concept, [12] proposed RTTE, a novel real-time travel time estimation framework with sparse surveillance information, showcasing its potential as a practical solution to real-time traffic monitoring and estimation in urban areas.

2) Bus Travel Time Estimation: Bus travel time estimation differs from that of cars due to the nature of fixed routes and designated stops, which has different travel patterns, more complex bus network representation reflecting important travel locations (e.g., stops, intersections), and fewer datasets. In the domain of public transport systems, the estimation of travel time for buses is widely recognized as a crucial yet challenging task. As noted previously, [4] constituted a commonly acknowledged baseline approach, displaying noteworthy performance results. Reference [16], on the other hand, integrated traffic forecasting information into [4], resulting in a proposed model which achieves a 21% improvement in accuracy when compared to the existing method without traffic predictions for bus travel time estimation.

Petersen *et al.* [17] utilized deep neural networks, specifically a combination of convolutional and long short-term memory (LSTM) layers, to effectively capture the non-static spatio-temporal correlations for bus travel time estimation. Pang *et al.* [7] approached the problem as a Multi-Step-Ahead Prediction (MSAP) task and offered a framework consisting of Recurrent Neural Network (RNN) with LSTM to leverage the long-range dependencies for prediction. Nonetheless, the methods mentioned above are not sufficiently comprehensive to learn from the spatial-temporal relationships. To be specific, [17] focuses on capturing temporal dependencies alone and neglects the spatial factors. Besides, in [7] only spatial dependencies between stops or intersections on a single route are considered. The interdependence of buses with similar routes is ignored. To address this issue, Ma *et al.* [18] proposed MAGTTE, an end-to-end multi-attention graph neural network-based model that demonstrates exceptional robustness when estimating bus travel times with highly sparse travel records.

Despite the promising domain of bus travel time estimation, this work places greater emphasis on estimating travel time for cars in the subsequent sections, which provides more generality to the context.

B. Route Representation

As an important part of TTE modeling, an appropriate route representation helps to mitigate errors and enhance the precision of predictions. Existing efforts on estimating travel time generally fall into three categories from the route representation perspective: segment-based, path-based, and OD-based methods.

1) Segment-based: The principle of "divide and conquer" forms the basis for segment-based methods, wherein a given path is decomposed into a sequential arrangement of road segments. The overall travel time is then calculated as the sum of each segment's time. Early results that rely on static sensors (e.g., loop detectors [19], [20]) are regarded as typical segment-based methods which possess the merit of low computation cost and high comprehensibility in line with intuition. However, such approaches suffer from the apparent drawbacks of disregarding delays from complex traffic conditions such as intersections and traffic lights. Additionally, they fail to uncover correlations among road segments; for instance, congestion on one particular segment may impact the travel times of adjacent ones.

To address the aforementioned challenges, authors in [9] utilized dynamic programming to estimate the travel time of an entire path, i.e., to search for the optimal combination of sub-path comprised of multiple road segments, each with pre-determined travel times. Despite that time delays can be incorporated into the estimation, there remains residual time spent at sub-path junctions overlooked.

Tang *et al.* [21] proposed a tensor-based spatial-temporal model to estimate the travel time of all road segments on a city scale under varying traffic conditions at different times of the day. This approach incorporates the spatial correlation between road segments and latent regularity captured from traffic condition fluctuation. Zygouras *et al.* [22] predicted the travel time for each individual road segment by utilizing information obtained from other segments with similar traffic characteristics, which is then summed up to calculate the travel time for a given query path. Qiu *et al.* [23] developed an innovative neighbor-based deep learning method to leverage the historical data at fine-grained time intervals and the features of adjacent segments, leading to improved TTE precision for each road segment.

In summary, segment-based methods exhibit several inherent shortcomings: a) The dynamic and uncertain nature of transportation systems makes it challenging to accurately predict road segment traffic conditions. As a result, estimation accuracy for each segment is partially affected. b) The errors stemming from the earlier-mentioned matter tend to amass, thereby resulting in a considerable estimation bias. Besides, ignoring intersection and traffic light delays also results in inaccurate time prediction of the entire path. c) Impact of personalized driving on travel time is neglected when considering paths as sequences of road segments. For instance, drivers residing nearby typically drive faster within the local region than those visiting an unfamiliar area. Thus, explicit differences exist between the travel times. d) Trajectory data must be projected onto the road network, which entails computationally heavy map-matching.

2) Path-based: Owing to the impressive ability to distill valid information and intrinsic connections of deep learning techniques, deficits of segment-based methods can be partially resolved by path-based end-to-end models. These models enable the capturing of complex traffic conditions and correlations implicitly [4], [24], [25]. Moreover, taking into consideration the path level data instead of relying on sequences of segments acquired through map-matching algorithms results in a reduction in computational expenses and eradication of errors arising from the algorithms.

DeepTTE [4] takes the raw GPS sequence as input and estimates the travel time of the whole path directly. Specifically, it employs sliding windows, each containing several consecutive sampled points, to transform a path into a sequence of windows. However, the performance of DeepTTE highly relies on both the accuracy and sampling frequency of raw GPS data. When intermediate points between the origin and the destination are unavailable, its performance rapidly declines.

Zhang *et al.* [24] took further steps to partition the road network into equal-sized grids before converting a trajectory composed of GPS-sampled points into a sequence of grids that it passes through. Nonetheless, the difficulty of defining the appropriate granularity of grid cells limits its further performance. Coarse-grained cells aggregate multiple sampled points within a cell, losing the precise movement information. Overly small cell sizes, on the other hand, result in a large number of grids with few or almost no points falling in. Under the circumstances, the correlation between similar trajectories is diluted, which in turn leads to severe data sparsity problems.

Drawing inspiration from natural language processing

(NLP), Wide-Deep-Recurrent (WDR) [25] was proposed as an innovative deep learning framework that simultaneously trains wide linear models, deep neural networks, and recurrent neural networks. In this model, each route is treated as a sentence and each road segment, along with its corresponding interaction, is considered as a word. By leveraging both the wide and the deep models to capture overall statistical properties of routes along with the use of recurrent models to capture detailed characteristics of road segments, both route- and segment-level features are utilized for accurate predictions.

Furthermore, Fu *et al.* [26] employed a sliding window over a given path to plot sub-paths in each window as images, treating paths as sequences of images to seamlessly capture spatial and temporal patterns.

It is worth noting that the prevalence of data-driven methods and machine learning tools has resulted in a blurring of boundaries between segment-based and path/sub-path-based approaches. Han *et al.* [27] exploited the advantages of different route representations and proposed a multi-semantic model for travel time estimation (STTE) to construct multiple informative representations for a path. This is achieved by considering the path both as a sequence of segments and as a sequence of intersections.

Additionally, there is also a growing trend to introduce multi-task learning frameworks during the training phase in some learning-based models [11], [13], [26], which aims to balance the trade-off between individual segment time estimation and overall path time estimation.

3) OD-based: In many real-world online services, only the origin and destination of a trip are provided, rather than the actual route before making a travel time estimation and it's time-consuming and error-prone to infer the likely route in advance. Furthermore, due to privacy preservation policies and tracking costs, access to entire trajectory information is often limited. Given these considerations, performing effective origin-destination (OD) travel time estimation assumes significant importance.

A nearest neighbor based method was proposed in [2] wherein the authors estimated OD travel time with the utilization of historical trips with similar origin, destination, and time of day. This approach avoids expensive route computation and achieves outstanding performance in terms of running speed (i.e., up to 40 times faster than other route-based methods at the time), which is essential for online services. However, as trip length increases, there may not be sufficient neighboring trips available (i.e., it is faced with severe data sparsity), thereby reducing confidence in travel time estimates.

In addition to the heuristic design, recent advances in addressing the OD travel time estimation problem have employed deep learning-based techniques. For instance, MURAT [14] is a multi-task representation learning model that leverages topological structure and spatio-temporal prior knowledge of road networks for OD travel time estimation. The proposed multi-task learning framework enhances learning performance by capturing meaningful path information from historical trips as an auxiliary task. Yuan *et al.* [28] proposed an effective encoding model that first generates spatio-temporal representations for historical trajectories and then binds the OD input to its affiliated trajectory during training phases. During prediction phases, only the OD input is used to generate its representation akin to generating a proper trajectory, resulting in improved estimations.

IV. Data-driven Techniques Perspective

A. Tensor-based Models

A tensor is a high-dimensional generalization of vectors and matrices in the form of multi-dimensional arrays of numerical data. Tensors can be used to represent the multivariate relationship between heterogeneous data, making them a natural choice for modeling variables in transportation systems with high-order tensors for the problem of TTE. For example, [9] constructs a third-order tensor where each element denotes the travel time of a particular driver on a specific road at a certain time. However, many road segments are not traversed during a non-negligible amount of time slots rendering a large volume of values in the tensor missing. To address this issue, tensor decomposition, an effective and versatile technique for dimension reduction, sparse data filling, and implicit relationship mining, is widely used to alleviate data sparsity in tensor-based TTE methods.

Specifically, authors in [9] extracted three categories of features, consisting of geospatial, temporal, and historical contexts. Then extracted features were fused with tensor decomposition to infer the missing values in tensors, namely the context-aware tensor decomposition approach. Similarly, Tang *et al.* [21] proposed a novel algorithm named *Probabilistic Traffic Condition Clustering*, which models travel time and its corresponding occurrence probability based on traffic conditions using two three-order tensors. They also introduced a context-aware tensor decomposition approach named *wCPr*, which accurately estimates missing entries. Huang *et al.* [29] took congestion level into consideration and incorporated it into their tensor model as a third dimension. Additionally, they proposed a coupled tensor decomposition algorithm that utilizes POI features to enhance the accuracy of missing data recovery.

In addition to the aforementioned tensor-based models, there are also studies that combine tensors with learning-based models to contribute towards feature extraction. For instance, Shen *et al.* [3] leveraged non-negative tensor decomposition to restore travel speed distributions within the Travel Speed Features Layer of their model where a Convolutional Neural Network (CNN)-RNN model is also integrated to extract both long-term and short-term travel speed features.

B. Learning-based Models

1) Deep Neural Network Models:

a) CNNs & RNNs: Due to their powerful ability for image processing, CNNs have a wide application on face recognition, object detection, and medical image analysis. For TTE, paths can also be projected into Euclidean space, thereby images. Inspired by this idea, CNNs are introduced for travel time estimation to exploit its outstanding power by capturing spatial correlations from image data.

For instance, Fu *et al.* [26], as discussed in Section III-B.2, leveraged a CNN-based approach to extract spatial motion patterns from sequences of sub-path images. To be specific, a unique two-dimensional CNN architecture, named PathCNN, was introduced to incorporate diverse pooling techniques for handling heterogeneous information within the images. Moreover, its convolution was also regulated for better capturing spatial features of lines instead of image textures as classical CNN models do. It is noteworthy that temporal dependency was also captured by the one-dimensional CNN model, which exhibits novelty compared with the prevailing usage of RNN and its variants for extracting temporal features. In [30], Lan *et al.* first divided the high-resolution map into a number of equal-sized grids and integrated trajectory data with it. A deep CNN was then employed to recognize patterns from morphological layout images of traversed grids and generate an effective representation for subsequent predictions.

RNNs are highly suitable for modeling temporal data owing to their memorization capability. Their variants, such as Long Short-Term Memory (LSTM) [18], [31], [32] and Gate Recurrent Unit (GRU) [13], find extensive application in capturing long-term temporal dependencies for TTE. Moreover, bidirectional LSTM (BiLSTM) is also adopted to enhance the LSTM by utilizing backward information [24], [30], [33]. Zhang *et al.* [24] designed a dual interval loss mechanism in the prediction layer for auxiliary supervision to further optimize the BiLSTM model's performance, which aligns perfectly with the characteristics of BiLSTMs and leads to better predictions.

b) GNN: Graph Neural Networks (GNNs) are a type of neural network that directly perform spectral convolution operations or apply spatial attention on graph structures with the ability to capture complex spatial dependencies between nodes and links. GNNs are currently regarded as state-of-the-art techniques for traffic forecasting problems including TTE and are ideally suited to these problems, as a road network is naturally a graph. For instance, Google Maps deployed a GNN-based estimator for TTE in production, resulting in significant reductions in negative results across different regions worldwide (e.g., 40+% reduction in Sydney) [34].

A heterogeneous information network (HIN) was introduced to the TTE task in HetETA [35]. A double-stuffed sandwich structure was designed wherein two GNNs are placed between three CNNs. GNNs were employed to encode spatially diverse information from HINs, while CNNs were utilized to process temporal information. Moreover, empirical findings validated that the learned representations by HetETA can be integrated as additional features into WDR, a meticulously designed feature system for TTE on Didi Chuxing's platform, leading to enhanced performance. Jin *et al.* [36] proposed STGNN-TTE based on the core architecture of ST-GCN [37], a deep learning framework based on Graph Convolutional Networks (GCNs) and temporal CNNs, which was initially proposed for traffic flow prediction. By consolidating the output of the designed multi-scale ST-GCN and extended temporal dynamics extracted by the Transformer layer, real-time traffic condition representation was learned, leading to a more refined estimation.

Also, the Graph Attention Network (GAT), a novel graph neural network architecture with an attention mechanism, is employed to spatial-temporal tasks. Fang *et al.* [11] pointed out that the majority of studies adopting spatial-temporal graph neural networks make use of spatial and temporal information separately, which neglects their joint relations. To address this issue, an elaborate spatial-temporal graph attention network, 3DGAT, was proposed in [11] to fully exploit the joint relations of spatial and temporal information and exhibited potential applications to other spatial-temporal problems as well. Fu *et al.* [38] focused on improving the inference speed of TTE and developed a learning system called CompactETA. In this approach, a GAT is applied to the road network graph to learn the spatial dependency between roads, while positional encoding techniques [39] are utilized to embed temporal dependencies. Specifically, it provides an accurate travel time prediction within $100\,\mathrm{ms}$ reducing the inference time by more than 100 times over other algorithms.

2) Learning Techniques:

a) Federated learning: In light of growing concerns regarding data security, conventional centralized-training models that necessitate the collection of trajectory data closely linked to the personal location of all individuals are increasingly vulnerable to privacy breaches. Consequently, there is a pressing need for novel TTE methods that do not require data sharing and prioritize protecting privacy. Federated learning is a distributed machine learning technique with a decentralized architecture that enables client-side data retention while facilitating collective model training on the server side, which is well-suited for TTE with privacy-preserving.

Given the decentralized and area-based deployment of traffic management systems, it is imperative to consider privacy-preserving data exchange when conducting TTE research that involves trajectory data covering multiple administrative areas within a city. To this end, Zhu *et al.* [40] designed a comprehensive cross-area privacy-preserving solution incorporating federated learning, which enables them to train a tailor-made travel time estimator in each area by local data while maintaining strict privacy protections.

In traditional federated learning schemes, all participants ultimately utilize the same model, thereby limiting its performance across different clients. To address this problem, Zhang *et al.* [41] proposed a federated learning system, GOF-TTE, compromising both a base global model and a fine-tuned personalized model. Benefiting from personalized tweaks and its real-time perception of the global traffic state, it makes more accurate predictions with the consideration of privacy issues.

b) Meta-Learning: Meta-learning is a learning paradigm that aims to acquire general knowledge across diverse tasks and subsequently transfer this knowledge to

TABLE I
OPEN DATA OF TTE

Dataset Name	Article
Porto Taxi[1]	[24] [26] [30] [22] [23] [13] [36] [27] [43] [44]
Chengdu Taxi	[4] [26] [13] [36] [27] [41] [40] [44]
NYC Taxi[2]	[2] [14]
T-Drive trajectory[3]	[33]
Beijing taxi[4]	[9] [21]

novel tasks, thereby achieving swift adaptation with minimal training data. Meta-learning algorithms are particularly well-suited for scenarios where data is scarce or rapidly changing, as they can learn from data without requiring extensive manual engineering. Notably, recent meta-learning algorithms employed in TTE can be classified into two categories: model-based [42] and optimization-based [43], [44].

Drawing on prior research conducted on Baidu Maps [11], Fang *et al.* [42] made notable advancements in the realm of prediction accuracy by incorporating the traveled route (i.e., the route already traversed from the origin to the driver's current location) into their analysis. This enables them to model driving preferences more effectively and address a novel task, namely *en route travel time estimation* (ER-TTE). To overcome the challenge of few-shot learning inherent in ER-TTE, a model-based meta-learning approach known as SSML was proposed, which leverages the limited observed driving behaviors to acquire meta-knowledge that enables rapid adaptation to a user's driving preference.

Nonetheless, it is widely acknowledged that model-based methods generally exhibit weaker generalization capabilities than their optimization-based counterparts. Specifically, Fan *et al.* [43] pointed out that trajectories with different contextual information tend to have different characteristics in ER-TTE so that directly using the same model for all trajectories is prone to incur inaccuracy as SSML does. Hence, they proposed a novel framework based on MAML [45], one of the most successful optimization-based meta-learning algorithms to date. Their approach offers personalized initial parameters and learning rates for each trajectory based on its specific contextual information. Also, Wang *et al.* [44] introduced optimization-based meta-learning techniques into the proposed MetaTTE framework to overcome challenges arising from dynamic temporal dependencies and changing road networks. This innovation opens up new avenues for providing continuously accurate travel time estimations over time for multi-city scenarios.

V. Experimental Data and Resources

In this section, we present a summary of the experimental data and source code utilized in the surveyed papers in Tables I and II. The data presented herein are particularly well-suited for conducting studies related to TTE, while the accompanying code resources serve as valuable tools for replicating previous TTE solutions as baselines in subsequent research endeavors.

TABLE II
SOURCE CODE OF TTE ALGORITHMS

Model Name	Year	Link
PTTE [9]	2014	`http://research.microsoft.com/apps/pubs/?id=217493`
DeepTTE [4]	2018	`https://github.com/UrbComp/DeepTTE`
ConSTGAT [11]	2020	`https://github.com/PaddlePaddle/Research/tree/master/ST_DM/KDD2020-ConSTGAT/`
HetETA [35]	2020	`https://github.com/didi/heteta`
TTPNet [3]	2020	`https://github.com/YibinShen/TTPNet`
SSML [42]	2021	`https://github.com/PaddlePaddle/Research/tree/master/ST_DM/KDD2021-SSML/`
MVSTM [31]	2021	`https://github.com/775269512/SIGSPATIAL-2021-GISCUP-4th-Solution`
MetaTTE [44]	2022	`https://github.com/morningstarwang/MetaTTE`

VI. Challenges and Future Directions

Despite notable advancements in existing studies, achieving even greater levels of accuracy in travel time estimation remains a significant goal for future research. In addition to this objective, this section also addresses four key challenges associated with TTE and presents several potential directions for future inquiry.

A. Heterogeneous Data

Achieving precise travel time estimation necessitates the incorporation of both spatio-temporal data and latent factors, such as weather conditions, driving styles, and points of interest (POI), among others. Consequently, heterogeneous data fusion represents a pervasive challenge in TTE research. Although significant progress has been made in fusing underlying graph structures with other information via GNNs, effectively generating appropriate representations for diverse types of heterogeneous data from disparate sources remains an unresolved issue. Early attempts have been made in this regard [10], [12], wherein inertial data collected from smartphones and surveillance information obtained from cameras are utilized respectively.

Furthermore, it should be noted that the current approaches are predominantly developed on trajectory datasets extracted from a single source. The reliability and precision of these techniques hinge heavily upon the availability of extensive data samples. However, this presents significant difficulties when dealing with certain categories of vehicles, such as public buses or ambulances, where obtaining large-scale data sets may prove to be an arduous task. In light of this

[1] `https://www.kaggle.com/datasets/crailtap/taxi-trajectory`
[2] `https://chriswhong.com/open-data/foil_nyc_taxi/`
[3] `https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/`
[4] `http://research.microsoft.com/apps/pubs/?id=217493`

challenge, authors in [32] proposed an alternative approach that utilizes trajectory data collected from heterogeneous vehicle sources in the same geographical region, which offers a promising solution to mitigate the aforementioned issue and unlock valuable insights for enhancing estimation accuracy.

B. Privacy-preserving

Given that trajectory data is inherently linked to location information, there exists a significant risk of exposing personal private information. In particular, the findings from de Montjoye *et al.* [46], who analyzed 15 months' worth of location data from 1.5 million individuals, revealed that as few as four space-time points are sufficient to uniquely identify up to 95% of these subjects. As such, with the growing recognition of data privacy concerns in contemporary society, it is imperative to address the challenge of preserving privacy during TTE estimation procedures.

In contemporary times, the acquisition of large-scale traffic data poses significant challenges. In many cases, released datasets may lack complete trajectory information and exclude intermediate GPS points. Moreover, stringent privacy regulations such as the General Data Protection Regulation [47] restrict service providers from sharing data with third-party entities to develop predictive models that require massive amounts of data, which limits the data acquisition when performing cross-area TTE as discussed in Section IV-B.2.a.

Therefore, advanced learning schemes, for instance, federated learning [40], [41], may be used to preserve users' data locally without data sharing in order to reduce the risk of privacy leakage. Moreover, various privacy-preserving techniques (e.g., Laplace differential private randomization mechanism [41] and geo-indistinguishability [48]) may be employed during data collection and publishing stages to safeguard against potential attacks and mitigate the risks posed by adversaries who have access to original datasets. Furthermore, there is a need for continuous research into developing robust estimator models that are capable of learning from limited information while preserving user privacy.

C. Uncertainty Estimation

The majority of existing TTE research endeavors are primarily focused on providing accurate and deterministic travel time estimations for given routes. However, the reliability of these estimations is often impacted by a multitude of dynamic factors (e.g., traffic conditions and human behaviors) that are difficult to record accurately. Therefore, uncertainty quantification should be incorporated into TTE models to predict the time uncertainties of trips and provide users with reachable time confidence estimates, which is crucial for making informed decisions regarding schedule planning and route selection. Notably, as demonstrated in [40], uncertainty models that estimate auxiliary arrival probabilistic distribution can describe TTE better than deterministic ones. To address the challenge, the Bayesian deep learning approach may be adopted, which can be viewed as a probabilistic extension to standard deep learning that enables quantification of both model and input uncertainty [49].

D. Different Travel Mode

The available trajectory datasets for TTE research are mostly collected from ride-hailing cars or buses, resulting in a narrow focus on the heterogeneous type of vehicles in current TTE research. However, travelers may utilize various modes of transportation, such as motorcycles and bicycles. In addition, they are liable to choose mixed-mode travel trips that entail cycling and public transportation, thereby necessitating the travel time estimator to recognize mode changes and adjust accordingly. Therefore, it is vital to develop adaptive estimators with the capacity to detect different travel modes from mixed-mode trajectory data, since it allows for the exploitation of heterogeneity, leading to more accurate estimations.

VII. Conclusion

This paper provides a comprehensive survey of data-driven methods for travel time estimation. Specifically, two key aspects of the TTE problem are examined in detail: application scenarios and route representation approaches. The data-driven techniques employed in existing results are also summarized, along with the latest collection of open datasets and code resources related to this topic. Moreover, we identify several challenges that must be addressed in future research endeavors and outline potential directions for further investigation.

References

[1] C. J. Messer, C. L. Dudek, and J. D. Friebele, "Method for predicting travel time and other operational measures in real-time during freeway incident conditions," *Highway research record*, vol. 461, pp. 1–16, 1973.

[2] H. Wang, X. Tang, Y.-H. Kuo *et al.*, "A simple baseline for travel time estimation using large-scale trip data," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, Jan 2019.

[3] Y. Shen, C. Jin, J. Hua, and D. Huang, "TTPNet: A neural network for travel time prediction based on tensor decomposition and graph embedding," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 9, pp. 4514–4526, 2022.

[4] W. Dong, J. Zhang, W. Cao *et al.*, "When will you arrive? Estimating travel time based on deep neural networks," in *Proceedings of 32nd AAAI Conference on Artificial Intelligence*, New Orleans, Louisiana, USA, 2018, p. 2500–2507.

[5] T. Reich, M. Budka, D. Robbins, and D. Hulbert, "Survey of ETA prediction methods in public transport networks," *CoRR*, vol. abs/1904.05037, 2019.

[6] A. Miglani and N. Kumar, "Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges," *Vehicular Communications*, vol. 20, p. 100184, 2019.

[7] J. Pang, J. Huang, Y. Du *et al.*, "Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3283–3293, 2019.

[8] C. Whong. Foiling nyc's taxi trip data. (2014). [Online]. Available: http://chriswhong.com/open-data/foil_nyc_taxi/

[9] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, New York, USA, 2014, p. 25–34.

[10] R. Gao, X. Guo, F. Sun *et al.*, "Aggressive driving saves more time? Multi-task learning for customized travel time estimation," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, Macao, China, 2019, pp. 1689–1696.

[11] X. Fang, J. Huang, F. Wang *et al.*, "ConSTGAT: Contextual spatial-temporal graph attention network for travel time estimation at baidu maps," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Virtual Event, CA, USA, 2020, pp. 2697–2705.

[12] W. Zhang, Y. Wang, X. Xie *et al.*, "Real-time travel time estimation with sparse reliable surveillance information," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 1, Mar 2020.

[13] G. Jin, H. Yan, F. Li *et al.*, "Spatial-temporal dual graph neural networks for travel time estimation," *CoRR*, vol. abs/2105.13591, 2021.

[14] Y. Li, K. Fu, Z. Wang *et al.*, "Multi-task representation learning for travel time estimation," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, United Kingdom, 2018, pp. 1695–1704.

[15] Y. Sun, K. Fu, Z. Wang *et al.*, "CoDriver ETA: Combine driver information in estimated time of arrival by driving style learning auxiliary task," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 5, pp. 4037–4048, 2022.

[16] L. Tran, M. Y. Mun, M. Lim *et al.*, "DeepTRANS: A deep learning system for public bus travel time estimation using traffic forecasting," *Proc. VLDB Endow.*, vol. 13, no. 12, pp. 2957–2960, Aug 2020.

[17] N. C. Petersen, F. Rodrigues, and F. C. Pereira, "Multi-output bus travel time prediction with convolutional LSTM neural network," *Expert Systems with Applications*, vol. 120, pp. 426–435, 2019.

[18] J. Ma, J. Chan, S. Rajasegarar, and C. Leckie, "Multi-attention graph neural networks for city-wide bus travel time estimation using limited data," *Expert Systems with Applications*, vol. 202, p. 117057, 2022.

[19] Z. Jia, C. Chen, B. Coifman, and P. Varaiya, "The PeMS algorithms for accurate, real-time estimates of g-factors and speeds from single-loop detectors," in *Proc. 2001 IEEE Intelligent Transportation Systems*, 2001, pp. 536–541.

[20] K. F. Petty, P. Bickel, M. Ostland *et al.*, "Accurate estimation of travel times from single-loop detectors," *Transportation Research Part A: Policy and Practice*, vol. 32, no. 1, pp. 1–17, 1998.

[21] K. Tang, S. Chen, and Z. Liu, "Citywide spatial-temporal travel time estimation using big and sparse trajectories," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 12, pp. 4023–4034, 2018.

[22] N. Zygouras, N. Panagiotou, Y. Li *et al.*, "HTTE: A hybrid technique for travel time estimation in sparse data environments," in *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Chicago, IL, USA, 2019, pp. 99–108.

[23] J. Qiu, L. Du, D. Zhang *et al.*, "Nei-TTE: Intelligent traffic time estimation based on fine-grained time derivation of road segments for smart city," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2659–2666, 2020.

[24] H. Zhang, H. Wu, W. Sun, and B. Zheng, "Deeptravel: A neural network based travel time estimation model with auxiliary supervision," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, 2018, pp. 3655–3661.

[25] Z. Wang, K. Fu, and J. Ye, "Learning to estimate the travel time," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, United Kingdom, 2018, pp. 858–866.

[26] T.-y. Fu and W.-C. Lee, "DeepIST: Deep image-based spatio-temporal network for travel time estimation," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, China, 2019, pp. 69–78.

[27] L. Han, B. Du, J. Lin *et al.*, "Multi-semantic path representation learning for travel time estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 13 108–13 117, 2022.

[28] H. Yuan, G. Li, Z. Bao, and L. Feng, "Effective travel time estimation: When historical trajectories over road networks matter," in *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, Portland, OR, USA, 2020, pp. 2135–2149.

[29] L. Huang, Y. Yang, H. Chen *et al.*, "Context-aware road travel time estimation by coupled tensor decomposition based on trajectory data," *Knowledge-Based Systems*, vol. 245, p. 108596, 2022.

[30] W. Lan, Y. Xu, and B. Zhao, "Travel time estimation without road networks: An urban morphological layout representation approach," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, Macao, China, 2019, pp. 1772–1778.

[31] Z. Liu, Z. Wu, M. Wang, and R. Zhang, "Multi-view spatial-temporal model for travel time estimation," in *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*, Beijing, China, 2021, pp. 646–649.

[32] X. Lin, Y. Wang, X. Xiao *et al.*, "Path travel time estimation using attribute-related hybrid trajectories network," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, China, 2019, pp. 1973–1982.

[33] F. Wan, L. Li, K. Wang *et al.*, "MTTPRE: A multi-scale spatial-temporal model for travel time prediction," in *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, Seattle, Washington, 2022.

[34] A. Derrow-Pinion, J. She, D. Wong *et al.*, "ETA prediction with graph neural networks in google maps," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Virtual Event, Queensland, Australia, 2021, pp. 3767–3776.

[35] H. Hong, Y. Lin, X. Yang *et al.*, "HetETA: Heterogeneous information network embedding for estimating time of arrival," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Virtual Event, CA, USA, 2020, pp. 2444–2454.

[36] G. Jin, M. Wang, J. Zhang *et al.*, "STGNN-TTE: Travel time estimation via spatial–temporal graph neural network," *Future Generation Computer Systems*, vol. 126, pp. 70–81, 2022.

[37] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, 2018, pp. 3634–3640.

[38] K. Fu, F. Meng, J. Ye, and Z. Wang, "CompactETA: A fast inference system for travel time prediction," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Virtual Event, CA, USA, 2020, pp. 3337–3345.

[39] A. Vaswani, N. Shazeer, N. Parmar *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio *et al.*, Eds., vol. 30, 2017.

[40] Y. Zhu, Y. Ye, Y. Liu, and J. J. Q. Yu, "Cross-area travel time uncertainty estimation from trajectory data: A federated learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24 966–24 978, 2022.

[41] Z. Zhang, H. Wang, Z. Fan *et al.*, "GOF-TTE: Generative online federated learning framework for travel time estimation," *IEEE Internet of Things Journal*, vol. 9, no. 23, pp. 24 107–24 121, 2022.

[42] X. Fang, J. Huang, F. Wang *et al.*, "SSML: Self-supervised meta-learner for en route travel time estimation at baidu maps," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Virtual Event, Singapore, 2021, pp. 2840–2848.

[43] Y. Fan, J. Xu, R. Zhou *et al.*, "MetaER-TTE: An adaptive meta-learning model for en route travel time estimation," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, Vienna, Austria, Jul. 2022, pp. 2023–2029.

[44] C. Wang, F. Zhao, H. Zhang *et al.*, "Fine-grained trajectory-based travel time estimation for multi-city scenarios based on deep meta-learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15 716–15 728, 2022.

[45] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, vol. 70, Sydney, NSW, Australia, Aug. 2017, p. 1126–1135.

[46] Y.-A. de Montjoye, C. A. Hidalgo, C. A. Hidalgo *et al.*, "Unique in the Crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, p. 1376, 2013.

[47] C. of European Union. Regulation (eu) 2016/679 of the european parliament and of the council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). european commission. (2016). [Online]. Available: https://gdpr-info.eu/

[48] F. Liu, D. Wang, and Z. Xu, "Privacy-preserving travel time prediction with uncertainty using GPS trace data," *IEEE Transactions on Mobile Computing*, vol. 22, no. 1, pp. 417–428, 2023.

[49] F. De Sousa Ribeiro, F. Calivá, M. Swainson *et al.*, "Deep bayesian self-training." *Neural Computing & Applications*, vol. 32, no. 9, pp. 4275–4291, 2020.