*Journal of Internet Services and Applications, 2024, 15:1,*doi: 10.5753/jisa.2024.4041

Evaluation of Trajectory and Destination Prediction Models: A Systematic Classification and Analysis of Methodologies and Recent Results

João Batista Firmino Júnior [**Instituto Federal de Educação da Paraíba** | ***batista.firmino@academico.ifpb.edu.br***]
Janderson Ferreira Dutra [**Instituto Federal de Educação da Paraíba** | ***janderson.dutra@ifpb.edu.br***]
Francisco Dantas Nobre Neto [**Instituto Federal de Educação da Paraíba** | ***dantas.nobre@ifpb.edu.br***]

Instituto Federal de Educação da Paraíba, Av. Primeiro de Maio, 720, Jaguaribe, João Pessoa, PB, 58015-435, Brazil

Abstract
Predicting trajectories and destinations is of considerable relevance in the context of urban mobility, as it can be useful for suggesting detours, avoiding congestion, and optimizing people's commutes. Therefore, this research performs a classification and analysis of trajectory and destination prediction models in articles published from 2017 to 2023. These models were mapped considering: authors; the existence of more than one geographic scenario; the type of forecast; the use of semantic and contextual data; and description of the algorithms. The result consists of discussions of representative works, based on classification, with grouping of techniques. Furthermore, there is a focus on works that used contextual and/or semantic data, from which another framework was developed, specifying the titles of the articles, and whether the methodology involved the use of points or areas of interest, and a reference to how they were generated. This focus expands the previous framework, specifying the differences of a portion of published studies.

Keywords: Trajectories, Destinations, Predictions, Semantic Data, Contextual Data

1 Introduction

In recent years, many studies on trajectory and destination prediction have been carried out with the aim of defining a specific method or algorithm that improves prediction accuracy. Studies focused on predictive models have produced varied results, which serve as a basis for future investigations and identification of challenges inherent to this field of study.

Given the variability of these models, their approaches, their results and challenges, it was necessary, for the 2017 to 2022 period - together with some 2023 onwards articles - to carry out a Systematic Literature Mapping (SLM). And in this space, there is the search for a mapping which can describe and analyzing certain techniques, methods, and algorithms, to provide support for new research.

Aiming for an updated overview of the most recent research in a specific area of knowledge, some tools for literature review and analysis are essential. Among the possible tools, this SLM - based on the article originally published by Firmino Júnior *et al*. (2023) -, sought to classify and analyze works inherent to the state of the art in the researched theme.

With this purpose, the analysis was partially based on the protocol defined by Kitchenham and Charters (2007), which includes a detailed plan for the review, specifying the process to be followed and the conditions for its application in the selection of primary studies. Considering that the research scope in this SLM is focused on the prediction of trajectories and/or destinations, with objectively defined criteria for selecting literature articles, the main contribution of this study is to highlight models that provide a path for addressing the following Research Questions: "How were the trajectory and/or destination prediction models constructed from 2017 to 2023?" e "What are the challenges brought, based on the differences in these works, to the continuity of research in the area?".

The specific contributions of this SLM, therefore, are:

- A comparative and detailed table of 33 works selected after adopting an adaptation of the methodology established by Kitchenham and Charters (2007);
- A comparative analysis of each of the methodological differences and results;
- The outline of some challenges for the continuity of basic research in the area, anchored in comparative analysis;
- Emphasis on the most used recent techniques;
- Focus on works that explored semantic and/or contextual data. As a challenge to be highlighted, it is necessary to find ways through interpretation of works that used these types of data.

In addition to this introductory section, this article is structured as follows: Section 2 presents the guiding Basic Concepts; Section 3 covers Related Works, encompassing the evolution of this field of study based on the most prominent works; Section 4 highlights the methodological path defined for the SLM; in Section 5, the results obtained are explained, with a comparative table of the selected studies and a detailed explanation of the methodological differentiators, with subsection 5.1 focusing on works with semantic and/or contextual data; In Section 6 - the conclusion - the consequences of Section 5 are discussed in terms of the challenges encountered with 6.1 subsection.

2 Basic Concepts

This section introduces the main concepts associated with the lines of research with trajectory prediction and destination prediction.

- Trajectories: For Leite da Silva *et al*. (2019), correspond to a series of points in chronological order, knowing that they may contain other information such as categories of places visited - in the latter case, called "multiple aspect trajectory". According to Graser *et al*. (2023), these data can address various domains such as Computer Science, Geography, Urban Planning, and Ecology. Predicting trajectories means anticipating patterns of movement of an object or user. According to Liu *et al*. (2019), considering a moving object, the trajectory is a finite sequence, a sequential arrangement of space-time points sampled from tracking through some location device. Such prediction may involve data associated with the purpose of each route, such as semantic data according to Sadri *et al*. (2018).
- Destinations: Because of the definition by Leite da Silva *et al*. (2019), the intended destinations are understood from these trajectories. Destinations should not be confused with intermediate stopping points, which are points where people stay for some time before reaching their intended location according to Zheng (2015). The activity of predicting destinations, consequently, is about finding the most likely destination based on historically defined patterns. These patterns can stem from trajectories.
- Spatial Data: Considering trajectories, this concept refers to computational representations in which attributes are computationally processed and stored with their respective geometries, through a Geographic Information System (GIS) according Druck *et al*. (2004). This means dependence on the presence of geographic regions.
- Temporal Data: This concept consists of data of the datetime type, which is revealed through the date format plus hours, minutes and seconds. They can have a specific zone with its own time count or be Universal Time Coordinate (UTC) or Global Time Greenwich Mean Time (GMT), which are international standards that apply to the entire planet. The first, UTC, was initially created in the early 1960s to improve the dissemination of an earlier system, UT1. It uses Global Navigation Satellite Systems (GNSS), according to Arias and Guinot (2004). The second, GMT, means solar time and, according to the Royal Greenwich Observatory, is older than UTC, but should not be used for more precise purposes, according to Weintrit (2017). In summary: these are data that bring the idea of continuity of the phenomenon over a given space.
- Contextual Data: Consists of data common to all users, and relates to the context or location in which the events are located, such as weather data, topographic data (in relation to the terrain), traffic signs and day of the week. Liu *et al*. (2019) they consider, for example, the number of taxis requested, travel demand (the Local Spatial Context), whether all districts are residential (which the authors call the Global Relational Context) and meteorological data (highlighted as the Temporal Evolution Context).
- Semantic Data: Data relating to the function of a place, such as a destination. It is observed that a location can represent different semantics for users with different profiles. An example consists of capturing movement points and stop points for semantic enrichment, making it possible to infer the motivation for these different states in which the moving objects are found according to Santana and Campos (2017).

Thus, considering the aforementioned concepts, the methodology was developed. However, it is necessary to position the relevant works in the field, considering the last decades, and their relationship with this SLM.

3 Research Trends in Trajectory and Destination Predictions

The studies on the prediction of space-time data for car moving are grounded in older technologies and techniques. Among the technologies, there is the Global Positioning System (GPS), which provides information such as longitude, latitude with date and time. The GPS allows bringing, with specific accuracy, a succession of points along a route, making it possible to generate polyline geometry, which can be approximated to real paths through Map-matching techniques (with the aim of refining the location of the polyline of the trajectory of urban vehicles with that of geographic features, such as roads) or with the adequacy between different Coordinate Reference Systems.

Furthermore, much is owed to Machine Learning algorithms and the techniques and tools of Geographic Information Systems (GISs, are systems that involve hardware, software, methodologies, geographic data and qualified professionals), that were boosted by advances in computing in recent decades.

The relevance of this lies in the complexity of technological artifacts, algorithms, systems, and methodologies, besides concepts of Smart Cities, which involves a whole other field of knowledge with a separate chapter for Transport - in addition to the field called Big Data. What materializes in a State of the Art with some studies that stood out in recent decades.

3.1 Main Approaches

There are three types of approaches, in terms of techniques used, for predicting trajectories and/or destinations: (i) predictions involving Machine Learning (such as Random Forest, Support Vector Machine, and Decision Tree, for example) with algorithms that are relatively less complex than techniques with Deep Machine Learning. This set of works is less numerous, but it exists; (ii) approach in several works with a highlight on the use of Markovian Models - either individually or in conjunction with other techniques; (iii) work with models that use Deep Learning techniques, such as LSTM, CNN and others.

Considering the categories defined above, the first relevant study was that of Froehlich and Krumm (2008), which used Hidden Markov Models. In the model developed by the researchers, trajectory repetitions are emphasized, revealing routes and reducing uncertainties in predictions through tests on a real dataset.

Regarding destination prediction, Krumm and Horvitz (2006) had already worked with driver behavior data, using 40 by 40 m cells, each of them representing a discretized place. They used probabilistic methods, considering unvisited locations and the frequency of visitation and types of visited locations. The tests were conducted in three scenarios, and the results were evaluated with Mean Average Precision (MAP).

Starting from the last destination visited, Gambs *et al*. (2012) considered Markov Chains to predict new destinations, and, for two previously visited locations, the accuracy was 70 percent to 95 percent.

Another relevant study, but this one done with Hidden Markov Models, was carried out by Simmons *et al*. (2006), who predicted routes and destinations by distributing continuous data in boxes. In terms of contextual data, they considered the day of the week, time of day and speed.

When it comes to the category of works with relatively simpler Machine Learning algorithms, those by Zhang *et al*. (2020) and Araújo *et al*. (2019). In the first one, the model was developed with Support Vector Regression (SVR) with and without a technique called Circular Fuzzy Embedding (CFE), evaluated by MAE and RMS. The second brings a combination of Markov Model with Random Forest and was evaluated using Accuracy and F1-score.

Finally, in relation to the category of works with Deep Learning algorithms, the works related by Graser *et al*. (2023), especially those that used Long-Short Term Memory (LSTM), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Transformers, Convolutional Neural Network (CNN) and Self-Attention Network (SAN).

In common, they are models developed according to different technical perspectives and varied methodologies, but their objectives are centered on predicting trajectories and/or destinations. However, they depend on available datasets, specific computational capacity, and the experience of previous related works, which are constantly advancing.

Thus, there is a predominance in the use of Markov Models and Deep Learning. It is also necessary to add that only in this millennium has there been a more prominent advance in access to public datasets, as well as to certain algorithms.

4 Methodology

This research considered studies published from 2017 to 2023, which met the Research Questions mentioned in the Introduction section. The protocol adopted included three main steps defined by Kitchenham and Charters (2007), as shown in Figure 1.

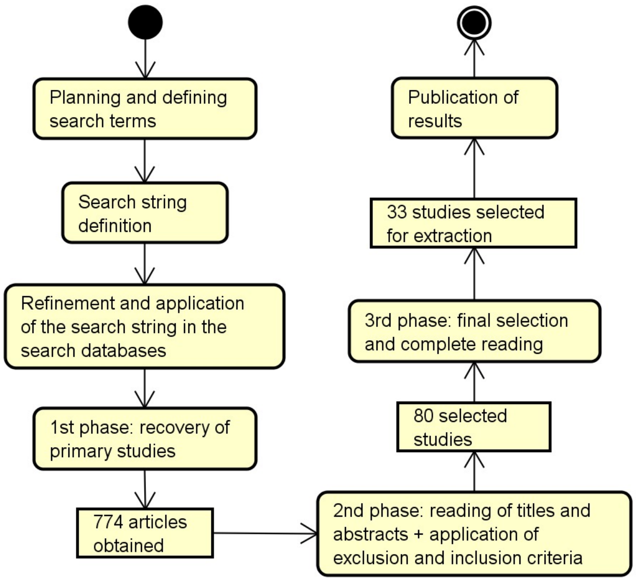

Figure 1. General search steps.

In the planning stage, the protocol and objective of the work were defined. Initially, the following search bases were considered: Association for Computing Machinery(ACM) Digital Library, Institute of Electrical and Electronics(IEEE) Xplore, Tandfonline, GEOINFO, IJCAI-17 and the search engine of the Brazilian Computing Society (SBC). As for SBC articles, the search was based on terms such as “routes”, “route prediction”, “routes” and “route prediction”. For the other bases, the search string was applied: (”trajectory prediction” OR ”destination prediction” OR ”route planning”) AND ”roads”.

The adopted protocol underwent initial tests, leading to slight changes in descriptors and search strings. It was necessary to adapt the search string across databases for calibration to achieve better results. Thus, initially, 794 studies were obtained. In the conduction stage, the retrieved primary studies were read and evaluated, initially, by titles, keywords, abstracts, introductions and principles. Finally, a complete reading of the studies was carried out, looking for a possible correlation of patterns between these recent works.

The inclusion criteria for selecting articles were: (i) being in English or Portuguese; (ii) be a complete study (avoiding articles with solutions still in progress); and (iii) have been published up to 5 years before 2022; (iv) use of urban means of transport such as bicycles, cars and buses, or walking. Excluded studies: (i) with an indirect approach to the topic of this article, such as speed prediction, focused more on route planning without including prediction of trajectories or destinations; (ii) studies outside the topic, such as networks and the Internet, or addressing certain moving objects outside the urban or road context, such as planes and boats; (iii) referring to autonomous vehicles or whose technique or object of investigation covered the theme of computer vision.

In addition to the selection criteria, the exclusion of studies that did not develop models for predicting trajectories and/or destinations was considered. For example, in cases involving computer vision or a focus on autonomous vehicles without human discretion. The quality criteria considered were presence of real datasets to validate the solutions proposed in the works; access to scripts, and datasets with already standardized data. Data were extracted, synthesized and organized in an electronic spreadsheet . Finally, the SLM results were described and evaluated.

In the end, 33 articles were defined to describe and evaluate the results.

5 Results

Through the readings and analyzes carried out, we came to a summary of what was found with the classification and analyzes.

The classification follows as in Table 1.

This table has six columns, where: the first contains the authors of the analyzed works; the second indicates whether there was more than one geographical scenario, with yes (S) or no (N); the third contemplates the type of prediction carried out in the work, in which the value "C" represents collective prediction, and the value "I" represents individual prediction, and (I and C) for both; the fourth shows whether (S) or not (N) the use of semantic data in predictive models; and the fifth column provides the context of the prediction made, in which the value "U" represents that the predictions were made in an urban environment, and "O" represents "Other" types of places. The last column describes the algorithms used in each article or the models developed.

Vahedian *et al*. (2017), Liu *et al*. (2019), Besse *et al*. (2018), Jiang *et al*. (2022), Ebel *et al*. (2020), Lassoued *et al*. (2017), Bhuvaneswari *et al*. (2017) and Tong *et al*. (2021) present models that, in this section, serve to highlight destination prediction. This does not eliminate the fact that there are articles that address both trajectories and destinations, in the context of route planning, or not.

The model proposed by Wang *et al*. (2017), focusing on destinations, aims to examine changes in the distances of these destinations found, in trajectory queries, instead of relying on historical trajectories, and on a sparse dataset; its main technique being Mobility Gradient-based Destination Prediction, with statistical evaluation metrics such as variance and information entropy. All this also with semantic data.

While the model by Imai *et al*. (2018) is in this classification, as it seeks to create a model to improve the accuracy in predicting new destinations from the initial stage of a trip. But it also seeks to improve the robustness of trajectory prediction. The proposed model works both in predicting destinations and trajectories, using techniques that involve trajectory tracking and NPP (Next Place Prediction).

Finally, they validated their results with precision measures, in addition to the percentile of the top-k correct destinations; then, they compared the adopted baseline with a multiclass logistic regression in combination with naïve bayes.

Vahedian *et al*. (2017), for example, stand out for presenting a model that aims to predict future collective events by forecasting/ predicting destinations of incomplete trajectories. The technique used is Via-Location Grouping, which validates the results through destination prediction accuracy and comparisons of memory cost and processing speed.

As for Liu *et al*. (2019), the proposed model is the demand for taxis between pairs of regions in a future time interval, involving destination prediction. It employs Neural Networks, and the results are validated through Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

Regarding the need to use trajectory data to predict destinations, there is the work of Besse *et al*. (2018), which is based on the partial initial trajectory. The model uses Gaussian distribution as its main technique. The results are presented using the Receiver Operating Characteristic (ROC) curve, the Area Under Curve (AUC), in addition to Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Jiang et al. (2022), sought to predict destinations based on a probabilistic method called P3M. Its main innovation is in the user behavior model called "Walker-Riding-Walking Probabiblistic Trip". The results are statistically evaluated using the Mean Absolute Error (MAE) metric.

Ebel *et al*. (2020) predict likely destinations and routes of a vehicle. To do this, they are based on the most recent partial trajectory and, as techniques, they use LSTM and Multilayer Perceptron (MLP), with their statistical evaluation metric being Accuracy, with datasets from the city of Porto (Portugal) and São Francisco (United States). Basically, differences were made between predicted values and real values, which were the trigonometric relationships between partial trajectories and centroid coordinates. The authors emphasize that such destination prediction is equal to the coordinates of the centroids of the candidate destinations closest to the extension of the straight line between the first and last point of the partial trajectory.

Lassoued *et al*. (2017), Bhuvaneswari *et al*. (2017) and Tong *et al*. (2021), respectively, deal with: presenting a model and a new algorithm for predicting destinations and routes with minimal complexity, with K-Means and hierarchical clustering with Markov Models. Thus, they generate statistically evaluated results with a Precision that consists of considering destinations and routes, and the average number of links. They propose a solution to first find the driver's destination and then choose the shortest route to reach that destination, with Hidden Markov Model (HMM), which evaluates results with accuracy statistics and average hop count. They develop a graph-based framework that considers carpooling locations for all students with real-world constraints, with DBSCAN, Decision Tree and Map-Matching, evaluating through metrics such as Precision, Recall and F1-Score).

However, being more specific in terms of predicting trajectories, or even adding to the prediction of destinations and trajectories, there is another set of works, sometimes coinciding with those in this subsection or not. In the next subsection, we will see the models of authors who focused more on trajectory prediction or included trajectory prediction in their models that already covered destination prediction.

As for articles focused on trajectory prediction, although there is some information about destinations, the works of Sadri *et al*. (2018), Chen *et al*. (2019), Ma and Xie (2021), Fu and Lee (2020), Barth *et al*. (2020), Tang *et al*. (2021), Liang and Zhao (2022), Dai *et al*. (2019), Rainbow *et al*. (2021), Fan and Yao (2017), Zhang *et al*. (2018), Wu *et al*. (2020), Qiao *et al*. (2018), Bhuvaneswari *et al*. (2017), Selvaraj (2021), Choi *et al*. (2019), Yuan and Li (2019), Tong *et al*. (2021), Chang *et al*. (2022), Santana and Campos (2017) and Araújo *et al*. (2019).

Chang *et al*. (2022) show a method that aims to complete

Table 1. Classification of models used.

Authors	More than 1 Scenario	Prediction type	Semantic Data	Context	Algorithms or Models
Wang et al., 2017	S	C	S	U	Training and prediction algorithms
Imai et al., 2018	N	I	N	U	Clustering Algorithms
Vahedian et al., 2017	N	C	N	U	Learning and Clustering Algorithms
Sadri et al., 2018	S	I	N	U	PreHeat and the “TrAf”
Chen et al., 2019	N	C	N	U	Long-short Term Memory (LSTM)
Ma and Xie, 2021	N	C	N	U	FCM (Fuzzy C-means) and LSTM
Fu and Lee, 2020	S	I	N	U	Recurrent Neural Networks (RNN) and Gradient Descent
Barth et al., 2020	S	I and C	S	U	Personalized Path Trajectory Segmentation (PPTS) and Optimal Path Trajectory Segmentation (OPTS)
Liu et al., 2019	N	C	S	U	Deep Neural Network Algorithms
Tang et al., 2021	N	C	S	U	Ridesharing Group Discovery and P-PPM Destination Prediction
Liang and Zhao, 2022	S	I	N	U	Generating Output Trajectory and LSTM
Dai et al., 2019	N	I	N	O	LSTM
Rainbow et al., 2021	S	I	S	U	Deep Neural Networks Algorithm
Besse et al., 2018	S	C	S	U	Hierarchical Clustering
Jiang et al., 2022	N	I	S	U	Decision Assistent and Crowd-Sourced Rebalancing
Fan and Yao, 2017	N	I	N	U	Transfer map into fixed points and Road-Based Location Sign (RBLS)
Ning, 2021	N	C	N	U	LSTM
Zhang et al., 2018	N	I	N	U	ESN, LSTM and Kalman Filter
Wu et al., 2020	N	I	S	U	Pedestrian Trajectory Prediction Algorithm
Ebel et al., 2020	S	I	N	O	LSTM
Lassoued, 2017	S	C	S	O	Cluster Prediction
Qiao et al., 2018	S	C	N	O	PrefixTP
Bhuvaneswari et al., 2017	S	C	N	O	SAHDID Prediction Algorithm
Selvaraj et al., 2021	S	I	N	U	LSTM
Choi et al., 2019	S	I	N	U	Feed-Foward Neural Network (FFNN)
Yuan and Li, 2019	N	I	N	O	DISON
Tong et al., 2021	N	I	N	U	Dijkstra Algorithm, Tabu-Based Expansion and Greedy Expansion Algorithm
Ren et al., 2022	S	I	S	U	DBSCAN, Brooks-Iyengar (BI) Algorithm and K-means
Santana and Campos, 2017	N	I	S	O	Algorithm that groups points, or isolated points
Araújo et al., 2019	S	C	S	U	TEMMUS
Qin et al., 2023	N	C	S	U	Urban topology-encoding spatiotemporal attention network (UTA)
Schen et al., 2023	S	I	N	U	Spatio-Temporal Interactive Graph Convolutional Network (STI-GCN) with Gated Recurrent Unit (GRU) in the construction of graphs and Convolutional Neural Network (CNN) for temporal features
Wang et al., 2023	S	I	S	U	Lane Transformer

a user's daily trajectory using the current day's trajectory in addition to historical paths, based on the Markov model and LSTM, and being validated by RMSE and p-value. The authors Chen *et al*. (2019) attempt to predict urban taxi trajectories by proposing a model with Neural Networks. Its main technique is LSTM, and validation occurred with BLEU Score.

Ma and Xie (2021), the model's contribution consists of predicting not exactly the trajectory of a vehicle, but of surrounding vehicles. The model adopted Fuzzy C-Means and LSTM techniques. And validated the work with MAE and RMSE.

Fu and Lee (2020) propose a new framework called Trembr (although the focus was on trajectory prediction), used for a series of applications with trajectory data. Its main techniques involve proprietary methods with RNN (Recurrent Neural Network). Validation occurred through MAE between the predicted result and the real value.

An article with semantic data, as in Wang *et al*. (2017), was developed by Wang *et al*. (2017), which aimed to provide a semantic understanding of collected trajectories - however in Sadri *et al*. (2018), it is not clear whether prefix patterns indicate semantic information. In any case, the proposed model works with multi-criteria segmentation of trajectories and validates the work with Segment ability Score, Break Recovery State, Segmentation Rate and Segmentation Quality Score.

More focused on recommending shared trips based on GPS, also depending on trajectory prediction, is the model by Tang *et al*. (2021), whose main techniques were Prefix Prediction Partial Matching (P-PPM), Prediction Partial Matching (PPM) and Markov, and statistically evaluated by Prediction Accuracy.

Regarding the model known as NetTraj considers each trajectory as a sequence of intersections that are associated with the directions that a movement can take. NetTraj was presented and demonstrated in Liang and Zhao (2022), whose main technique was LSTM and the validation metrics were Distance Error (DE), Average Match Ratio (AMR) and Match Ratio k or MR(k).

Dai *et al*. (2019), in turn, proposes to predict trajectories in dense traffic. The Spatial-LSTM (ST-LSTM) technique was used and the statistical evaluation was carried out using the Mean Absolute Deviation (MAD), Root Mean Square (RMS) and MSE measurements. While Rainbow et al. (2021) proposed introducing class information into a Graph Neural Network (GNN) to better predict an individual's trajectory. The main technique was Semantics-STGCNN. And the work was evaluated by the quantitative metrics Average Displacement Error (aADE), Average Final Displacement Error (aFDE), and qualitatively by social and semantic data.

Fan and Yao (2017) propose a new approach to trajectory prediction with sparse data, similar to the model proposed in Wang et al. (2017) but focusing on trajectories. Their main technique was Road-based Location Sign (RBLS), and the work was statistically evaluated for accuracy.

In Ning (2021), there is the Spatio-Temporal Trajectory Model (STTM), using LSTM and with the metrics Mean Relative Percentage Error (MRPE), MAE, Mean Relative Error (MRE), RMSE and the coefficient of determination R2. The work aims to show, for example, that the climate has a relevant influence on predicting results, basically involving taxisy.

As for the work of Zhang *et al*. (2018), the proposed model aimed to fill the gap in predictions of very rarefied and short trajectories, using Neural Networks. Once again, the problem of sparse datasets is encountered. And in this case, the main techniques were RNN, LSTM and Echo State Network (ESN). Probability Density Function (PDF) and Cumulative Density Function (CDF) were the validation metrics.

Authors Wu *et al*. (2020) sought a method for predicting pedestrian trajectories considering the pedestrian's intention and behavior. Bayesian network and LSTM were used. And its validation metrics were Accuracy and Average Error.

Meanwhile, the paper by Qiao *et al*. (2018) presents PrefixTP, which is the implementation of a trajectory prediction algorithm based on prefix projection. His main techniques were prefix projection and grouping. The metric for validation was the prediction time.

Regarding the inclusion of trajectory prediction (along with destination prediction), in proposing a solution to first find the driver's destination and then choose a shorter route to reach that destination, there is a study by Bhuvaneswari et al. (2017), whose main technique was HMM, and the results were statistically evaluated by Accuracy and Average Hop Count.

In relation to the article by Selvaraj (2021), the prediction of trajectories at intersections is included, whose main technique was LSTM and the metric based on Cumulative Density Function, which implies yet another model based on Neural Networks.

For the model proposed by Choi *et al*. (2019), the prediction of arterial trajectories is proposed, which allows showing the next intersections that the vehicle will visit, based on stored data. The model considers each intersection as a point of interest and uses the Artificial Neural Network (ANN) technique for prediction. The results obtained were evaluated using the Accuracy measure.

Yuan and Li (2019), have, to support large-scale trajectory data, the Distributed In-Memory Trajectory Similarity Search and Join on Road Network (DISON), whose main technique was DISON with the Longest Common Road Segment (LCRS). The metrics used to evaluate the proposed model were User Study Accuracies, Noise-based Method and Clustering-based Method.

To find a way to extract 3D structures from grid-separated junctions of vehicle trajectories, in the context of trajectory prediction, Tong et al. (2021) used DBSCAN as the main technique. The metrics were True Positive for Slopes (TPS), False Positive for Slopes (FPS) and True Negative for Slopes (TNS).

In the article by Santana and Campos (2017), the objective is to, with semantic trajectories, propose a solution to reconstruct travel stories using data such as footprints, originating from online social sources. Its main technique was the modeling of data for an application, with no use of Machine Learning. Validation of the results was carried out by Accuracy.

Another article, by Araújo *et al*. (2019), introduces TEMMUS, a trajectory prediction model based on similarity

Markov models. Its main technique is a Markov model and its model evaluation metric is Accuracy.

Thus, with the descriptions of each work (and model), it is possible, in the Conclusion, to discuss what was done and the challenges that remain, as well as to refer to the predominance of types of techniques.

Qin *et al*. (2023) worked on predicting taxi trajectories in Hangzhou (China), considering semantic data in the development of the Urban topology-encoding spatiotemporal attention network model (which the authors call UTA).

For this model, the methodology was built with the characterization of a study area, moving on to the generation of semantic points (from areas or points of interest that were related to 20 categories, such as finance, transport, tourism, etc., within a radius of 1000 m from the centers of intersections between areas of interest and roads) and a topological map with types of roads, and the processing of trajectories (with more than 8 million points using electric or hybrid vehicles). From these steps it was possible to predict trajectories, with the encoding of the topological nodes included in the UTA model (which has a topological module and an attention calculation module), ranking of the vertices and evaluations using Area Under Curve (AUC) and Group Area Under Curve (GAUC), in addition to RMSE and MAE. Furthermore, they had as baseline HMM, LSTM, RNN, Graph Convolutional Network (GCN) and Transformer. The results of these evaluations were: RMSE/m and MAE/m respectively 600.28 and 179.46 (below all other cases), and a higher Accuracy than all baselines, at 99.4 percent. The improvement, with AUC and GAUC, was, respectively, 2.37 percent and 1.71 percent.

Shen *et al*. (2023) propose a Spatio-Temporal Interactive Graph Convolutional Network (STI-GCN), considering only spatio-temporal data. The model relies on constructing a spatial autocorrelation function to describe the degree of mutual influence between vehicles. It is evident that the granularity of the work is centered around the "vehicle" unit and its neighborhoods, considering spatial interactions. The methodological steps involve the model architecture, with the extraction of spatial features and the consideration of the relationship between spatial agents; they extract temporal features and use the negative log-likelihood minimum. These procedures involve data from the Next Generation Simulation (NGSIM) dataset and encompass baselines from other research by authors such as the Kalman Filter (for constant speed), o Vanilla-LSTM, ConvSocial-LSTM, Multi-Agent Tensor Fusion (MATF), GRIP (which is based on graphs and LSTM), SCALE-Net, Graph-Based Information Sharing Network (GISNet), and convolutional neural network model based on data segmentation method (DS-CNN). Performance was measured by RMSE. The authors considered different rates of missing data, per baseline, in addition to different parameter sizes and inference times, with STI-GCNN obtaining the values of 29.1K parameters in an inference time of 0.0321 Ms. In qualitative terms, in turn, they conclude that the STI-GCNN presents better results than those of the baselines, and that there is an important influence on the vehicles close to the one whose prediction is made. They provide the code: `https://github.com/Yutasq/Multi-Class-Social-STGCNN`.

Finally, Wang *et al*. (2023) propose the Lane Transformer model, which uses attention blocks instead of Graph Convolutional Networks, maintaining accuracy and reducing time cost. This prediction model was built considering compatibility with TensorRT. They used the Argoverse dataset, with an accuracy greater than 10 to 25 times compared to LaneGCN baseline, in addition to producing faster code, according to the authors. In addition to these elements, they provide the github link https://`https://github.com/mmdzb/Lane-Transformer` with the code. The metrics for evaluation were adapted according to the problem, namely the Minimum Average Displacement Error, the Minimum Final Displacement Error and the MR (Miss Rate). They were concerned with dividing the results into quantitative and qualitative aspects.

However, with all these descriptions and analyses, it is necessary to highlight a set of differences in these works, going beyond the inclusion and quality criteria defined in the Methodology of this SLM.

The differences are centered, in terms of usefulness for future work, on more contextual and semantic aspects of the data, considering part of the articles analyzed.

Contextual data, and mainly semantic data, involves quantifying a certain subjectivity in the user's perception of the environment that surrounds him/her.

5.1 Searches with Contextual and/or Semantic Data

Seventeen of the 33 works used, at some level, contextual and/or semantic data. In the table 2 below, the classification of this subsample, considering criteria such as: references, use of points of interest (PoIs) or Areas of Interest (AoI), whether Yes (Y) or No (N), and the Direct (D), Indirect (I) or Unknown (DESC) generation of contextual and/or semantic data. This can assist in the development of other research in terms of how to use data of this nature in work in this field.

This way, there is a detailed list of the use of this type of data, observing, without a doubt, the particularities and problems to be resolved by each proposal.

6 Conclusion

Even after considering, based on the methodological process, 33 works and models, refined with a subsample of 14 studies that used contextual and/or semantic data, it is still necessary to consider that the prediction of trajectories and destinations is a relatively recent field, which does not always have semantic data. It was then observed the need for metrics adapted to new techniques, algorithms, methods or frameworks. Furthermore, the absence of specialists in the field of Transport was noted, and no explicit observation about the field of Coordinate Reference Systems, so important in GIS, was made. Many of the works had a more computational perspective and less focused on Geographic Information Systems.

In general, there was a predominance of real data for validating and producing results; low frequency of direct availability of scripts or even low frequency of access to original

Table 2. Articles that used contextual or semantic data.

References	PoIs/AoIs	Generation
Araújo *et al.* (2019)	N	I (Foursquare)
Barth *et al.* (2020)	S	I (route preference)
Besse *et al.* (2018)	S	DESC
Jiang *et al.* (2022)	S	DESC
Lassoued *et al.* (2017)	S	D (OpenStreetMap)
Liu *et al.* (2019)	S	I (Didi Chuxing, Uber and Grab)
Qin *et al.* (2023)	S	D (surrounding space)
Rainbow *et al.* (2021)	N	D (by object type)
Chang *et al.* (2022)	N	D (considering altitude data)
Santana and Campos (2017)	N	I (through georeferencing)
Tang *et al.* (2021)	S	D (sparse data such as departure time)
Tong *et al.* (2021)	N	D (through travel profiles)
Wang *et al.* (2023)	N	I (through the number of road segments)
Wu *et al.* (2020)	N	I (environment around the pedestrian)

data, for reproducibility; little emphasis on how they worked with Coordinate Reference Systems and possible transformations between systems; and predominance of urban scenarios.

Thus, some research opportunities are approaching: the exploration of relevant trajectory information using GIS, and, depending on the maturation of these studies, a productive dialogue with specialists in Transport - especially, urban commuting -, and the consolidation of new validation metrics with statistics professionals.

6.1 Challenges

In terms of challenges, new models can be proposed, possibly with some emphasis on the use of GIS software and the logic of SRCs. Similarly, I focus on the use of Python libraries created in the last two years, such as Moving-Pandas and Scikit-Mobility, mainly in the set of steps that come before a prediction, which can be useful for Spatio-Temporal Data Mining.

What was stated in the previous paragraph reveals a need to better work on the steps prior to predicting trajectories and/or destinations. That is, trajectory mining, not to be confused with mining only spatio-temporal data but including contextual and semantic data.

This, however, depends on the configuration of each dataset. There are a variety of dictionaries for each purpose, without a standard. Defining a standard for a set of trajectory/destination data can be crucial in well-conducted research.

Other challenges are finding public datasets and in the case of predictions based on destinations or trajectories never visited, or traveled, before. Thus, the main challenges are presented. And, in future work, they can be better explored - especially this last one. However, it's possible infer a solution, as we will see below.

The challenges in this research area are provided in Rainbow *et al.* (2021). The authors observed that there are different patterns of trajectories, which influences the setup of experiments, and they emphasized that, in many studies, there are implicit correlations among the different types of road segments on each trajectory to be predicted are ignored. Therefore, they found that it is not as useful to use only relative distances. In this way, the authors considered the semantic aspect with an adjacent matrix of labels (Label Adjacency Matrix) combined with a VAM (Velocity-Based Adjacency Matrix), resulting the model named as SAM (Semantics-Guided Graph Adjacency). Their work divides the results in quantitative and qualitative aspects, makes use of Python and PyTorch library, publicizes the source code and the dataset, as shown in the Results section, which can be accessed at `https://cvgl.stanford.edu/projects/uav_data/`.Once again, there is an appropriate choice of dataset according to the research problem and its methodology. In summary, the authors first consulted existing works, thus considering eight frames as historical data, to predict the next 12 randomly sampled frames, where K = 20 from a predicted multinomial distribution. It was also noticed that the code had to be adapted to the problem, as well as the ADE (Average Displacement Error) and FDE (Final Displacement Error) metrics.

A possible solution for the dataset question, therefore, involves first thinking about the problem and associating it to another work. Next, comes to the choice of the dataset, which can be publicly available and for free. Regarding the computational cost, techniques can and should also be adapted to this aspect. And the programming language itself will only depend on the freedom to adapt the code to the problem. Regarding the metrics, they can also be adapted, as long as they can be explained reasonably easily. And, regarding the sampling, it is clear that different regions, or space-time windows, are normally considered as a certain number of times, in order to, in the end, compare different techniques with the one developed by the research. This appears to have been the case in the article previously explained.

Declarations

Acknowledgements

Thanks are due to the Postgraduate Program in Information Technology at the Federal Institute of Education of Paraíba (IFPB) and its Professional Master's Degree in Information Technology, and to all the professionals who form it.

Funding

There was no funding.

Authors' Contributions

JBFJr: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing original draft, Writing review & editing. JFD: Formal analysis, Investigation, Methodology, Visualization, Writing original draft, Writing review & editing. FDNN: Investigation, Methodology, Project administration, Supervision, Writing original draft, Writing review & editing.

Competing interests

There were no interest conflicts.

Availability of data and materials

`https://docs.google.com/spreadsheets/d/1z417YC284bzfDbWIW2jKiRJxMtz5H2l9XnheRqV5NZs/edit?usp=sharing`

References

Araújo, F. R. D., Rosário, D. L.; Machado, K., Cerqueira, E. C., and Villas, L. (2019). TEMMUS: A Mobility Predictor based on Temporal Markov Model with User Similarity. *Simpósio Brasileiro De Redes De Computadores E Sistemas Distribuídos (SBRC)*. DOI: 10.5753/sbrc.2019.7389.

Barth, F., Funke, S., Jepsen, T. S., and C., P. (2020). Scalable unsupervised multi-criteria trajectory segmentation and driving preference mining. In *Proceedings Of The 9th Acm Sigspatial International Workshop On Analytics For Big Geospatial Data*. DOI: 10.1145/3423336.3429348.

Besse, P. C., Guillouet, B., Loube, J.-M., and Royer, F. (2018). Destination prediction by trajectory distribution-based model. *IEEE Transactions On Intelligent Transportation Systems: A Publication Of The IEEE Intelligent Transportation Systems Council*, 19:2470–2481. DOI: 10.1109/TITS.2017.2749413.

Bhuvaneswari, M., Jyothirbindhu, V., and Paramasivan, B. (2017). Statistical adaptive heuristic model supported driver intended destination and path prediction algorithm for vehicular ad hoc network. In *2017 International Conference On Algorithms, Methodology, Models And Applications In Emerging Technologies (ICAMMAET)*, pages 1–6. DOI: 10.1109/ICAMMAET.2017.8186749.

Chang, R., Luliang, T., Xue, Y., and Jed, L. (2022). Mapping Grade-Separated Junctions in Detail Using Crowdsourced Trajectory Data. *IEEE Transactions On Intelligent Transportation Systems: A Publication Of The IEEE Intelligent Transportation Systems Council*, 23:5552–5561. DOI: 10.1109/TITS.2021.3054910.

Chen, Z., Liu, H., Wang, Y., Duan, Z., and Chen, C. (2019). E-dam: Encoder-decoder with attention mechanism for city-scale taxi trajectory prediction. In *Proceedings Of The 2019 3rd International Conference On Computer Science And Artificial Intelligence*. DOI: 10.1145/3374587.3374610.

Choi, S., Kim, J., Yu, H., and Yeo, H. (2019). Real-time prediction of arterial vehicle trajectories: Anapplication to predictive route guidance for an emergency vehicles. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 4030–4036. DOI: 10.1109/ITSC.2019.8917122.

Dai, S., Li, L., and Li, Z. (2019). Modeling vehicle interactions via modified LSTM models for trajectory prediction. *IEEE Access*, 7:38287–38296. DOI: 10.1109/ACCESS.2019.2907000.

Druck, S., Carvalho, M. S., Câmara, G., and Monteiro, A. V. M. (2004). Análise espacial de dados geográficos. Available at: `http://www.dpi.inpe.br/gilberto/livro/analise/`.

Ebel, P., Gol, I. E., Lingenfelder, C., and Vogelsang, A. (2020). Destination prediction based on partial trajectory data. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1149–1155. DOI: 10.48550/arXiv.2004.07473.

Fan, H. and Yao, W. (2017). A trajectory prediction method with sparsity data. In *2017 IEEE International Symposium On Parallel And Distributed Processing With Applications And 2017 IEEE International Conference On Ubiquitous Computing And Communications (ISPA/IUCC)*, pages 1261–1265. DOI: 10.1109/ISPA/IUCC.2017.00191.

Firmino Júnior, J. B., Dutra, J., and Nobre, F. (2023). Classificação e análise de modelos de predição de trajetórias e de destinos - um mapeamento sistemático da literatura. In *Anais do XV Simpósio Brasileiro de Computação Ubíqua e Pervasiva*, pages 41–50, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbcup.2023.229831.

Froehlich, J. and Krumm, J. (2008). Route prediction from trip observations. In *Society of Automotive Engineers (SAE) 2008 World Congress*. DOI: 10.4271/2008-01-0201.

Fu, T.-Y. and Lee, W.-C. (2020). Trembr: Exploring Road Networks for Trajectory Representation Learning. *ACM Transactions on Intelligent Systems and Technology*, 11:1–25. DOI: 10.1145/3361741.

Gambs, S., Killijian, M.-O., and del Prado Cortez, M. N. (2012). Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility - MPM '12*. DOI: 10.1145/2181196.2181199.

Graser, A., Jalali, A., Lampert J., Weißenfeld, A., and Janowicz, K. (2023). Deep Learning From Trajectory Data: a Review of Neural Networks and the Trajectory Data Representations to Train Them. *Workshop on Big Mobility Data Analysis BMDA2023 in conjunction with EDBT/ICDT 2023*. Available at: `https://ceur-ws.org/Vol-3379/BMDA_2023_paper_7556.pdf`.

Imai, R., Tsubouchi, K., Konishi, T., and Shimosaka, M. (2018). Early destination prediction with spatio-temporal user behavior patterns. In *Proceedings Of The AcmOn In-*

teractive, Mobile, Wearable And Ubiquitous Technologies, volume 1, pages 1–19. DOI: 10.1145/3161197.

Jiang, M., Li, C., Li, K., and Liu, H. (2022). Destination prediction based on virtual POI docks in dockless bike-sharing system. *IEEE Transactions On Intelligent Transportation Systems: A Publication Of The IEEE Intelligent Transportation Systems Council*, 23:2457–2470. DOI: 10.1109/TITS.2021.3099571.

Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Available at: `https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf`.

Krumm, J. and Horvitz, E. (2006). Inferring destinations from partial trajectories. In *UbiComp 2006: The Eighth International Conference on Ubiquitous Computing*. DOI: $10.1007/11853565_15$.

Lassoued, Y., Monteil, J., Gu, Y., G., R., R., S., and M, M. (2017). A hidden markov model for route and destination prediction. In *2017 IEEE 20th International Conference On Intelligent Transportation Systems (ITSC)*, pages 1–6. DOI: 10.1109/ITSC.2017.8317888.

Leite da Silva, C., May Petry, L., and V., B. (2019). A survey and comparison of trajectory classification methods. In IEEE, editor, *8TH Brazilian Conference on Intelligent Systems (BRACIS)*, pages 788–793, Salvador, Brazil. DOI: 10.1109/BRACIS.2019.00141.

Liang, Y. and Zhao, Z. (2022). NetTraj: A network-based vehicle trajectory prediction modelwith directional representation and spatiotemporal attention mechanisms. *IEEE Transactions On Intelligent Transportation Systems: A Publication Of the IEEE Intelligent Transportation Systems Council*, 23:14470–14481. DOI: 10.48550/arXiv.2106.11175.

Liu, L., Qiu, Z., Li, G., Wang, Q., Ouyang, W., and Lin, L. (2019). Contextualized spatial–temporal network for taxi origin-destination demand prediction. *IEEE Transactions on Intelligent Transportation Systems.*, 20:3875–3887. DOI: 10.48550/arXiv.1905.06335.

Ma, W. and Xie, H. (2021). Research on prediction algorithm of vehicle trajectory in front based on driving intention classification. In *2021 4TH International Conference on Data Science and Information Technology*. DOI: 10.1145/3478905.3478977.

Ning, F. (2021). Prediction and detection of urban trajectory using data mining and deep neural network. In *2021 International Conference on Big Data Analysis and Computer Science (BDACS)*, pages 39–46. DOI: 10.1109/BDACS53596.2021.00017.

Qiao, S., Han, N., J., W., Li, R. H., Gutierrez, L. A., and Wu, X. (2018). Predicting long-term trajectories of connected vehicles via the prefix-projection technique. *IEEE transactions on intelligent transportation systems*, 19:2305–2315. DOI: 10.1109/TITS.2017.2750075.

Qin, X., Zhiheng, L., Kai, Z., Feng, M., and Xin, J. (2023). Vehicle Trajectory Prediction via Urban Network Modeling. *Sensors*, 23. DOI: 10.3390/s23104893.

Rainbow, B. A., Men, Q., and Shum, H. P. H. (2021). Semantics-stgcnn: A semantics-guided spatial-temporal graph convolutional network for multi-class trajectory prediction. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2959–2966. DOI: 10.48550/arXiv.2108.04740.

Sadri, A., Salim, F. D., Ren, Y., Shao, W., Krumm, J. C., and Mascolo, C. (2018). "What will you do for the rest of the day?: An approach to continuous trajectory prediction. *Proceedings Of The Acm On Interactive, Mobile, Wearable And Ubiquitous Technologies*, 2:1–26. DOI: 10.1145/3287064.

Santana, A. V. and Campos, J. (2017). Travel history: Reconstructing semantic trajectories based on heterogeneous social tracks source. In ACM, editor, *Proceedings Of The 22nd Brazilian Symposium On Multimedia And The Web*, NY, USA. Available at: `https://sol.sbc.org.br/index.php/webmedia/article/view/5379`.

Selvaraj, D. C. e. a. (2021). Edge learning of vehicular trajectories at regulated intersections. In *2021 IEEE 94TH Vehicular Technology Conference (VTC2021-FALL)*, pages 1–7. DOI: 10.1109/VTC2021-Fall52928.2021.9625570.

Shen, G., Pengfei, L., Zhiyu, C., Yue, Y., and Xiangjie, K. (2023). Spatio-Temporal Interactive Graph Convolution Network for Vehicle Trajectory Prediction. *Internet of Things 24*. DOI: 10.1016/j.iot.2023.100935.

Simmons, R., Browning, B., Zhang, Y., and Sadekar, V. (2006). Learning to predict driver route and destination intent. In *IEEE Intelligent Transportation Systems Conference*. DOI: 10.1109/ITSC.2006.1706730.

Tang, L., Duan, Z., Zhu, Y., Ma, J., and Liu, Z. (2021). Recommendation for ridesharing groups through destination prediction on trajectory data. *IEEE Transactions On Intelligent Transportation Systems: A Publication Of The IEEE Intelligent Transportation Systems Council.*, 22:1320–1333. DOI: 10.1109/TITS.2019.2961170.

Tong, P., Du, W., Li, M., Huang, J., Wang, W., and Qin, Z. (2021). Last-mile school shuttle planning with crowdsensed student trajectories. *IEEE Transactions On Intelligent Transportation Systems: A Publication Of The IEEE Intelligent Transportation Systems Council*, 22:293–306. DOI: 10.1109/TITS.2019.2956786.

Vahedian, A., Zhou, X., Tong, L., and Li, Y.; Luo, J. (2017). Forecasting gathering events through continuous destination prediction on big trajectory data. In *Proceedings Of The 25th Acm Sigspatial International Conference On Advances In Geographic Information Systems.*. DOI: 10.1145/3139958.3140008.

Wang, L., Yu, Z., Guo, B., Ku, T., and YI, F. (2017). Moving destination prediction using sparse dataset: A mobility gradient descent approach. *ACM transactions on knowledge discovery from data*, 11:1–33. DOI: 10.1145/3051128.

Wang, Z., J, G., Ziqiang, H., Haiqiang, Z., Junping, Z., and Jian, P. (2023). Lane Transformer: A High-Efficiency Trajectory Prediction Model. *IEEE Open Journal of Intelligent Transportation Systems 4*, pages 2–13. DOI: 10.1109/OJITS.2023.3233952.

Weintrit, A. (2017). The concept of time in navigation. TransNav the International Journal on Marine Navigation and Safety of Sea Transportation. *TransNav*, 11:23–33.

DOI: 10.12716/1001.11.02.01.

Wu, H., Wang, L., Zheng, S., Xu, Q., and Wang, J. (2020). Crossing-road pedestrian trajectory prediction based on intention and behavior identification. In *2020 IEEE 23rd International Conference On Intelligent Transportation Systems (ITSC)*, pages 1–6. DOI: 10.1109/ITSC45102.2020.9294491.

Yuan, H. and Li, G. (2019). Distributed in-memory trajectory similarity search andjoin on road network. In *2019 Ieee 35th International Conference On Data Engineering (ICDE)*, page 1262–1273. DOI: 10.1109/ICDE.2019.00115.

Zhang, C., Bin, J., Wang, W., Peng, X., Wang, R., Halldearn, R., and Liu, Z. (2020). AIS Data Driven General Vessel Destination Prediction: A Random Forest Based Approach. *Transportation Research Part C: Emerging Technologies 118 (September)*. DOI: 10.1016/j.trc.2020.102729.

Zhang, W., Liu, Y., Liu, T., and Yang, C. (2018). Trajectory prediction with recurrent neural networks for predictive resource allocation. In *2018 14TH IEEE International Conference On Signal Processing (ICSP)*. DOI: 10.1109/ICSP.2018.8652460.

Zheng, Y. (2015). Trajectory Data Mining. *ACM Transactions on Intelligent Systems and Technology*, 6:1–41. DOI: 10.1007/978-3-319-17885-$1_1$576.