

Open Access

<https://doi.org/10.48130/DTS-2023-0014>
Digital Transportation and Safety 2023, 2(3):176–189

Taxi origin and destination demand prediction based on deep learning: a review

Dan Peng, Mingxia Huang* and Zhibo Xing

Department of Transportation and Geomatics Engineering, Shenyang Jianzhu University, Shenyang 110168, China

* Corresponding author, E-mail: mingxia@sjzu.edu.cn

Abstract

Taxi demand prediction is a crucial component of intelligent transportation system research. Compared to region-based demand prediction, origin-destination (OD) demand prediction has a wide range of potential applications, including real-time matching, idle vehicle allocation, ride-sharing services, and dynamic pricing, among others. However, because OD demand involves complex spatiotemporal dependence, research in this area has been limited thus far. In this paper, we first review existing research from four perspectives: topology construction, temporal and spatial feature processing, and other relevant factors. We then elaborate on the advantages and limitations of OD prediction methods based on deep learning architecture theory. Next, we discuss ongoing challenges in OD prediction, such as dynamics, spatiotemporal dependence, semantic differentiation, time window selection, and data sparsity problems, and summarize and compare potential solutions to each challenge. These findings offer valuable insights for model selection in OD demand prediction. Finally, we provide public datasets and open-source code, along with suggestions for future research directions.

Keywords: Deep learning; Taxi demand prediction; Taxi OD demand prediction; Spatiotemporal data mining; Dynamic graph

Citation: Peng D, Huang M, Xing Z. 2023. Taxi origin and destination demand prediction based on deep learning: a review. *Digital Transportation and Safety* 2(3):176–189 <https://doi.org/10.48130/DTS-2023-0014>

Introduction

As urban populations grow and motorization rates increase, the daily transportation needs of city residents have become more significant. However, the widespread use of private cars exacerbates traffic congestion, and large-scale public transportation systems are often limited in their ability to meet individualized travel needs due to issues such as coverage, operating hours, and route limitations. As a result, demand-responsive public transportation services such as taxis and ride-hailing have emerged as preferred modes of travel for city dwellers due to their high accessibility, all-day operation, and comfortable, quick services.

Since the issuance of the National Informatization Plan's '13th Five-Year Plan' by the State Council in 2016, intelligent transportation construction has become a significant focus in China's smart city development. Shared travel, with taxis and ride-hailing services playing a crucial role, has emerged as an important direction for this effort. Governments at provincial and municipal levels have released relevant planning documents to guide and support the development of various operating models, including new energy, ride-hailing, and cruise taxis, among others, with a focus on deep integration and intelligent services.

In recent years, problems associated with taxi services such as difficulties in accessing a taxi, long wait times, traffic congestion, and wastage of resources have become increasingly prominent. Accurate prediction of taxi demand can aid in rebalancing the spatial and temporal distribution of vehicle resources and alleviate the spatial and temporal discrepancies between the supply and demand of taxis.

The issues of taxi demand prediction include both node and edge forecasts. Node forecasts aim to predict the total number of trips for each region, while edge forecasts focus on predicting travel demand relationships between two regions

Currently, most research on taxi demand prediction focuses on forecasting the total passenger demand in a particular target area for a specific time frame. Deep learning techniques such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Graph Convolutional Networks (GCN), and their variations have been widely employed to extract temporal and spatial features for accurate predictions. In addition, several studies explore the incorporation of external factors, such as weather and points of interest (POI) in urban areas, to enhance prediction accuracy. Furthermore, researchers have utilized attention mechanisms, multi-task learning, residual networks, and other methods to further improve forecast accuracy.

Accurately predicting origin-destination demand is crucial for taxi platforms to make optimal real-time decisions regarding vehicle matching, idle vehicle reallocation, ride-sharing services, dynamic pricing, and other operational strategies. Origin-destination prediction involves forecasting the travel demand or origin-destination patterns of a particular region for a given period. OD demand prediction is more complex than regional-level demand forecasting due to its intricate spatial and temporal dependencies. However, given the current need to serve as many passengers as possible with limited taxi resources, OD demand prediction has a wide range of practical applications. Despite this, the existing research on OD demand prediction is limited. To address this gap, this paper aims to make the following contributions:

Review on taxi origin and destination prediction

This paper provides a systematic summary of existing research on taxi OD demand prediction, including methods used, challenges faced, and future research directions. The findings presented in this paper are intended to assist researchers in identifying areas for further investigation, as well as expanding existing research. Moreover, the practical applications of this research, which employs deep learning methods to enhance OD demand prediction, make this study highly relevant and timely. In conclusion, this paper aims to promote the application and development of OD demand prediction based on deep learning methods.

This paper provides a comprehensive review of the existing research on OD demand prediction that utilizes deep learning methods to process temporal and spatial features. The review delves into not only the theoretical aspects but also the advantages and limitations of these methods, aiming to inspire subsequent researchers to develop more novel models.

Furthermore, this paper discusses some of the key challenges that are faced by most OD demand prediction models. For each challenge, several existing solutions are summarized and compared, providing useful insights for selecting appropriate models in different contexts. Finally, based on the review and analysis conducted, this paper proposes future research directions for OD demand prediction.

This paper aims to facilitate baseline experiments in the field of transportation by collecting open-source datasets and codes from relevant literature. Additionally, this study proposes future research directions in the field.

Mathematical statistical methods

Statistical prediction methods are based on historical data and time series and belong to parameter methods. The commonly used models include the history average model, Auto-regressive Moving Average (ARMA) model, moving average (MA) model, auto-regressive integrated moving average (ARIMA) model, and Kalman filtering model.

Tebaldi & West^[1] used a Bayesian model to analyze the flow intensity between directed origin-destination (OD) pairs. Carvalho^[2] used a hierarchical Bayesian statistical model to address the problem of reconstructing static OD matrices. Spiess^[3] estimated the mean using a maximum likelihood model to estimate the OD matrix. Chang & Tao^[4] proposed a two-stage method for parallel computation that decomposed multiple subnets in the first stage and designed updated parameters for dynamic OD estimation in large-scale networks. They further developed a dynamic traffic assignment model for estimating time-varying network OD distributions. Chen et al.^[5] divided the uncertainty of estimating the OD matrix into two types: statistical uncertainty and the existence of multiple feasible OD demands on the same link. This was done to improve the prediction accuracy by determining the confidence interval. Hazelton^[6] proposed a Gaussian model based on the lower-level over-dispersed process and developed a Markov chain Monte Carlo algorithm for OD matrix prediction. Djukic et al.^[7] applied PCA to transform high-dimensional OD matrices into low-dimensional space and estimated real-time OD demand. Shao et al.^[8] proposed a heuristic iterative estimation allocation algorithm to optimize the path selection behavior for OD demand changes based on weighted least squares predictions of the mean and covariance matrix of OD demand. Lu et al.^[9]

proposed a dual-fluid curve analysis method and iterative matrix for dynamic OD route guidance. They calculated the dwell time based on iterative matrix calculations and conducted dynamic OD route guidance. Zhu & Guo^[10] proposed a LOESS method for urban event detection based on time series decomposition and anomaly detection at specific locations for OD big data urban event prediction. In the same year, Ren & Xie^[11] proposed a four-order tensor modeling method consisting of origin, destination, vehicle type, and time. By decomposing the tensor and extracting time factor matrices, it was used to predict future OD flow. However, the issue of data loss in high-dimensional data analysis remains unsolved. Li et al.^[12] proposed the NMF-AR method for OD matrix prediction through non-negative matrix decomposition (NMF) and auto-regressive (AR) modeling.

Although the statistical prediction methods based on mathematical statistics have made some progress by only extracting time-related features, their dimensionality is too simple. Taxi data is a typical spatiotemporal data set, and this method cannot extract spatial impacts, thus resulting in limited effect.

Traditional machine learning methods

Machine learning-based predictive methods belong to non-parametric methods, which are data-driven methods that can capture feature relationships in complex data. Commonly used methods include regression analysis represented by linear regression, Support Vector Machine (SVM), decision tree algorithm, Random Forest (RF), artificial neural network (ANN), and so on.

Support Vector Machine (SVM) is a non-linear regression method based on the minimization of risk structure criteria. When samples are linearly inseparable in the original space, SVM can use kernel functions to map samples from the original space to a high-dimensional space, making the samples linearly separable in the high-dimensional space. SVM can extract decisive features from small samples and is less prone to overfitting than other machine learning algorithms. However, when the sample size or number of dimensions is large, the model may run slowly and take longer.

Random Forest (RF) is composed of decision trees, which are common classification and regression algorithms. Decision trees consist of nodes and directed edges. At each internal node of each tree, the optimal feature for splitting is chosen according to a certain criterion, and the dataset is recursively divided into subsets. To avoid overly complex decision trees, pruning operations are performed. Random Forest uses two forms of randomness, sample Bagging and feature random subspaces, to learn from multiple decision trees and combines their results to make predictions for regression problems such as taxi demand prediction by taking the average of the decision tree results. Random Forest does not require pruning and is less likely to overfit, with good computational efficiency, robustness, and noise resistance.

Li et al.^[13] used the Quantum Particle Swarm Optimization (QPSO) algorithm to optimize the Radial Basis Function (RBF) neural network and established the QPSO_RBF neural network model to predict the demand for ride-hailing services in urban mixed areas, using passenger boarding demand, weather conditions, and road congestion ratio as input feature variables. Lu & Li^[14] compared historical averages, the ARIMA model, the KNN method, and the ANN model using Singapore taxi GPS

data and verified the superiority of ANN in long-term forecasting. Hong^[15] used a Support Vector Regression (SVR) model to predict future traffic flow and employed the Chaos Simulated Annealing (CSA) algorithm and seasonal index calculation method to measure the impact of periodic changes on future traffic flow. Tong et al.^[16] proposed the Lin-UOTD model, which aims to quickly adapt to changing application scenarios by using a simple machine learning model to predict future taxi demand.

The selection of features in machine learning-based predictive methods directly affects the accuracy of the prediction model. Compared with methods based on time series prediction, machine learning-based methods have certain improvements in accuracy and generalization ability. However, they have defects in processing high-dimensional data and cannot effectively solve the nonlinear correlation of complex multidimensional data.

The background of deep learning

Benefiting from the substantial usage of deep learning in the field of traffic prediction, taxi demand prediction has progressed from traditional time-series and machine learning models to utilizing deep learning frameworks such as CNN, RNN, GCN, and their variants to extract features from both temporal and spatial dimensions. Some studies have incorporated external data, such as weather and POI data in urban centers, to increase prediction accuracy. Additionally, some studies have applied attention mechanisms, multi-task learning, residual networks, and other methods to improve prediction accuracy.

To develop either node-level regional taxi demand prediction or origin-destination-based forecasting, historical order data with both temporal and spatial information need to be preprocessed. After considering the spatiotemporal characteristics and external factors, modeling and prediction can be conducted based on both temporal and spatial dimensions. Based on this analysis, we will elaborate on related work in four areas: spatial topology construction, spatial-dependent modeling, time-dependent modeling, and other factors.

Spatial topology construction

Various deep learning methods require different spatial topology construction and data mining tasks. For instance, traditional prediction methods based on statistical learning do not require spatial topology construction, while CNNs are designed to process raster data, and GCNs are usually utilized for processing graph data.

Raster

Convolutional neural network-based prediction typically involves using the raster method, in which the research area map is partitioned into grids of $H \times W$ and other sizes. To obtain the grid-to-grid OD matrix, the taxi flows within each divided grid are aggregated. However, transportation networks possess both spatiotemporal attributes and non-Euclidean structural characteristics, rendering the grid-based topology construction method inadequate for handling the non-Euclidean relationship of traffic data. In addition, the effectiveness of prediction is contingent on the rationality of the grid division. If the raster is too small, the same functional area may be split, resulting in a higher data volume that increases the difficulty of

prediction. Conversely, larger rasters make it challenging to extract demand features and reduce prediction accuracy.

Graph

The graph convolutional neural network is utilized for prediction, wherein travel demand data is transformed into images, i.e., non-Euclidean spatial data, to extract intricate spatial dependencies. Two graph construction methods exist: static and dynamic graphs. The first step is to construct a graph with the OD pairs serving as the nodes, and the features of the nodes and edges are included in the predictive network model.

Static graph
Static graph means that the modeling and representation learning of the graph by the model assumes that the graph structure is constant, and the construction method can be a distance measure and a Gaussian kernel function threshold to calculate the similarity between pairs of nodes to obtain an adjacency matrix, or directly use connectivity as different nodes to derive a binary adjacency matrix, in addition, some studies considering whether to add external features will also choose to build external information maps such as distance maps, traffic connectivity maps, semantic function maps, weather.

Dynamic graph
A dynamic graph can be categorized into two types: one in which nodes and edges continually change over time, and another where node and edge properties vary over time. Traditional graph representation learning frameworks generate static representations and overlook the dynamic nature of the content. As taxi demand exhibits a temporal-spatial dynamic dependency, it is essential to fully consider dynamic graph properties. The construction methods are classified into two categories: discrete-time dynamic graphs (DTDG) and continuous time dynamic graphs (CTDG)^[17,18].

(1) DTDG

The DTDG method defines a length, τ , and updates the embedding at each τ time unit. It constructs a dynamic adjacency matrix or a sequence of multiple graphs. Each graph represents a snapshot at one time step, which can be understood as a series of 'snapshots' of the changing graph. However, the DTDG method relies heavily on how to divide the time granularity. A coarse time granularity renders it difficult to perceive useful information such as trends, whereas a fine time granularity leads to excessive noise. Concerning OD demand prediction, the OD stream serves as a directed dynamic connection graph. Therefore, the DTDG method results in a loss of information due to the discrete segmentation of OD stream information.

(2) CTDG

The continuous time dynamic graph (CTDG) updates the node representation based on event data. The event typically includes the type of event, the location, and time of the event. Events such as crime, traffic accidents, and OD demand are updated by embedding this way. For instance, an OD request can be described using a tuple (e, l, t) , where e denotes the type of OD request, l represents the location, and t represents the timestamp. As the events appear sequentially rather than as snapshots, this method is more natural and practical for updating embeddings, such as DyRep^[19] and JODIE^[20]. However, the CTDG method can only capture the time dependence of finite time steps. Concerning OD demand prediction, due to the spatiotemporal demand imbalance, many OD pairs have no

Review on taxi origin and destination prediction

demand at a specific time. This results in a large number of zero values in some regions during certain periods, i.e., data sparsity.

Spatial dependency

Deep learning algorithms process spatial features through two main categories of convolutional neural networks (CNNs) and graph convolutional networks (GCNs). Transportation networks have both spatiotemporal attributes and belong to non-Euclidean structure networks. Traditional CNNs can only process Euclidean spatial data, whereas GCNs can process travel demand data into images and perform complex spatial dependence mining on non-Euclidean spatial data. Currently, most research is based on improving the performance of the model and enhancing the prediction level by changing its internal structure and adding external factors. This work focuses on the fundamental implementation principles of CNNs and GCNs, as well as their respective advantages and disadvantages.

Convolutional neural networks

Convolutional Neural Networks (CNNs) are deep feedforward neural networks based on convolution operations (Fig. 1). They take in two-dimensional matrices and extract local features from the convolutional kernels and matrices processed at the convolutional layer. The advantages of CNNs include feature selection, weight sharing, and pooling mechanisms. Firstly, the convolutional layer and pooling layer of CNNs can automatically extract spatiotemporal features of transportation networks, avoiding the difficulty of manually selecting features. Secondly, CNNs reduce the number of parameters that need to be trained through weight sharing, thereby reducing model complexity, the risk of overfitting, and improving model generalization ability. Finally, the pooling mechanism of CNNs reduces the number of neurons and improves the model's robustness to the invariance of input space translation, making

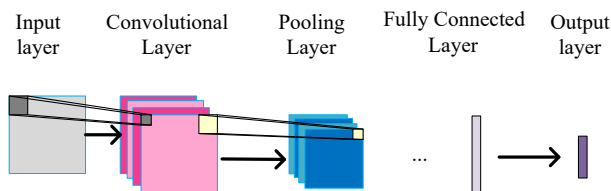


Fig. 1 Structure of convolutional neural network.

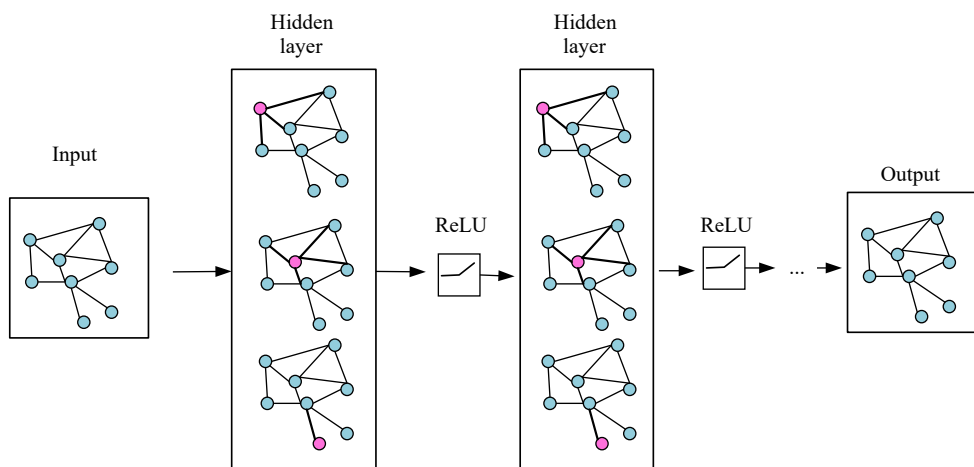


Fig. 2 Structure of graph convolutional neural network.

it suitable for large transportation networks and taxi demand prediction research^[21–32].

Graph convolutional neural networks

Although methods based on convolutional neural networks (CNN) can capture spatial correlations, they are best suited for processing spatial relationships in Euclidean space represented by two-dimensional matrices or raster images (Fig. 2). They lack the ability to handle data mining of non-Euclidean structures. In contrast, the graph neural network (GNN) contains a state variable that can represent any deep neighborhood information and captures the correlation of the graph structure through messaging between nodes. Therefore, GNN can meet the demand for taxi OD forecast on non-Euclidean space. Since its proposal in 2005, GNN has gradually been applied to taxi demand prediction models. GNN is divided into Spatial Convolution and Spectral Convolution based on different implementation methods. Spatial Convolution realizes the convolution operation through Graph Fourier Transform, and the processed graph structure must be an undirected graph. On the other hand, Spectral Convolution defines convolution as the aggregation of neighbor node features, which is more suitable for traffic road networks^[33–44].

Temporal dependency

Among the deep learning methods, there are three primary techniques for capturing time dependence: RNN and its variants, Transformer, and TCN. The cyclic operation of RNNs enhances model structure flexibility but at the cost of increased time and memory consumption. The introduction of GRU and LSTM has further improved the modeling ability of RNNs to depend on long-range sequences and avoid gradient disappearance. Transformer and its variations address the problem of long dependency by retaining the RNN's core functionality and adding an attention mechanism. However, this method can lead to quadratic computational complexity in long sequences. TCN utilizes parallelized causal and extended convolution to extract historical data features at distant moments while retaining long-term effective memory. Nevertheless, due to the small receptive field of TCN, it is not highly adaptable for transfer learning and is unsuitable for two-way learning.

Recurrent neural network

A recurrent neural network (RNN) is a neural network with a memory function capable of processing sequence data and

capturing relationships between sequences. During RNN's sequence data processing, the current time step input includes the input state x^t at the current moment t and the previous moment's output h^{t-1} to extract temporal features. However, when training the RNN network using backpropagation, it is vulnerable to the problems of gradient explosion and vanishing gradients. These issues become more noticeable with an increase in the number of cycles^[45-51] (Fig. 3).

The Long Short-Time Model (LSTM) improves upon the RNN model by introducing a gating mechanism that adds memory cell blocks in the hidden layer to resolve the problems of gradient explosion and vanishing. With its flexibility to adapt to the timing characteristics of various learning tasks, this method can capture deeper temporal characteristics. However, the high structural complexity of LSTM limits parallel computing, thereby prolonging model training time. Furthermore, traditional LSTM does not utilize spatial information encoded in input, which results in inadequate feature learning. Gated Recurrent Neural Network (GRU) simplifies the gating unit of the hidden layer and reduces computational cost while improving network computing power. Additionally, researchers have explored different LSTM architectures such as bidirectional LSTM and convolutional LSTM (Conv-LSTM). Conv-LSTM is a variant of LSTM that uses convolution operations instead of fully connected operations in input-state and state-state conversion, resolving the inadequacies of traditional LSTM^[52-54].

To simplify the gating unit of the hidden layer while improving network computing power and reducing time costs, the Gated Recurrent Neural Network (GRU) was developed based on LSTM. In addition, researchers have explored different architectures based on LSTM, such as bidirectional LSTM and convolutional LSTM (Conv-LSTM). Conv-LSTM is a variant of LSTM that uses convolution operations instead of fully connected operations in input-state and state-state conversion, resolving the inadequacies of traditional LSTM in utilizing spatial information encoded in the input^[55,56] (Fig. 4).

Transformer

The Transformer is the first transduction model relying entirely on self-attention to compute representations of its

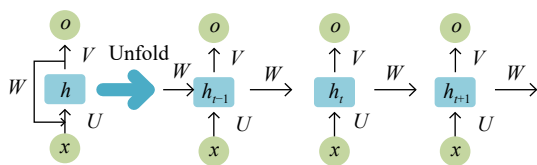


Fig. 3 Structure of recurrent neural network.

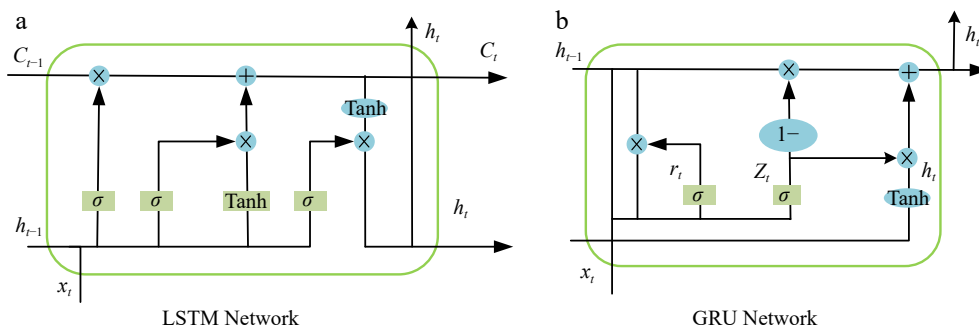


Fig. 4 Structure of LSTM network and GRU network.

input and output without using sequence aligned RNNs (Fig. 5). By using multi-head attention, the transformer solves the problem of long dependence, and also adopts parallel computing, residual connection, layer normalization, position coding, multi-head attention, and other technologies, so that the model has strong expression ability and computational efficiency, but the problem of quadratic computational complexity will occur in long sequences^[57,58].

Time convolutional networks

Time Convolutional Network (TCN) employs causal convolution and Dilated Convolution to permit parallelization (Fig. 6). Causal convolution is a one-way process that adheres to strict time constraints, with larger convolution kernels extracting increased historical information. Dilated Convolution introduces an input sequence expansion rate for controlling sampling intervals and extracting features from distant historical data for effective long-term memory retention and enhanced training outcomes. Despite these advantages, TCN's limited receptive field results in weak transfer learning adaptability and unsuitability for bidirectional learning^[59,60].

Other factors

External characteristics

Taxi demand is a time-evolving process influenced by external factors to a certain degree. Numerous studies have incorporated external factors such as weather^[61,62] and points of interest (POI)^[63,64] in data collection and preprocessing to aid prediction.

Model helper methods

Besides explicit external features, certain studies improve model performance by incorporating additional modules, such as the attention mechanism, multi-task learning, and ResNet network.

Attention mechanism

The attention mechanism was initially proposed in the seq2seq task to extract crucial information by filtering accepted data and appropriately allocating limited resources (Fig. 7). However, excessive application of the attention mechanism leads to increased computation time and memory demands. It slows down processing and is impractical for GPU training due to parallelization difficulties arising from the incorporation of the attention mechanism^[65,66].

ResNet network

Traditional convolutional neural network (CNN) structures face the issue of gradient disappearance and explosion when the depth of the network is increased. The ResNet network

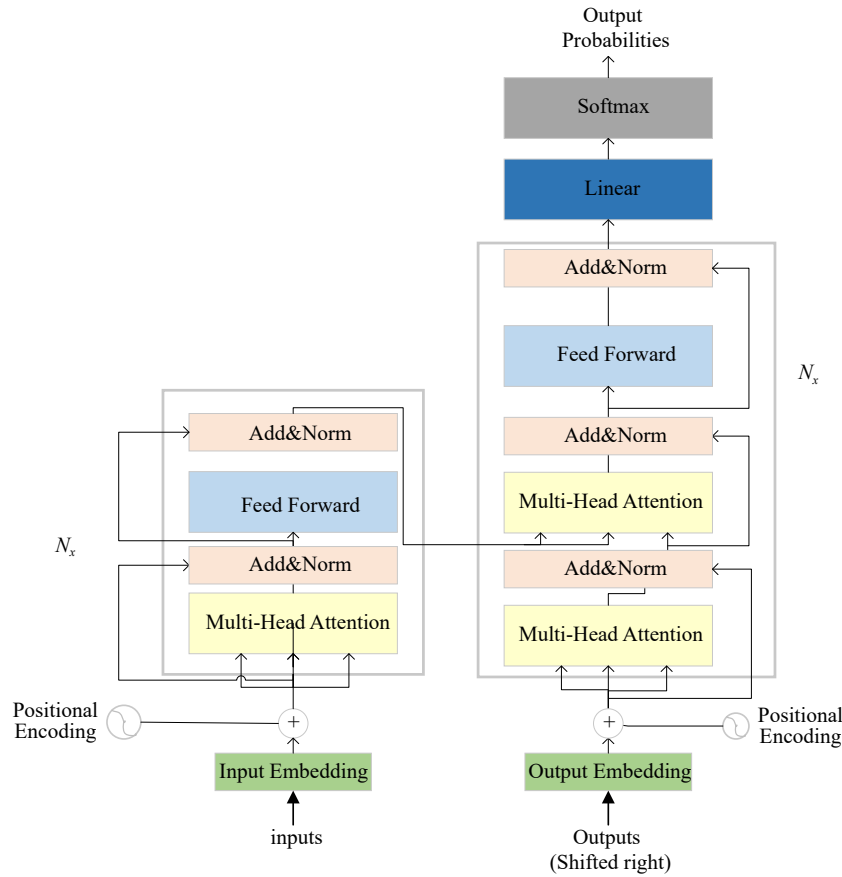


Fig. 5 Structure of Transformer network.

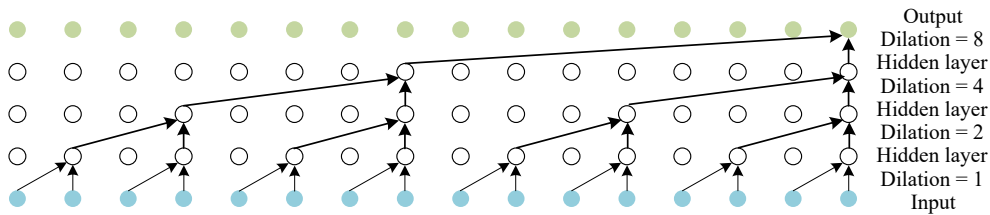


Fig. 6 Structure of Temporal Convolutional Network.

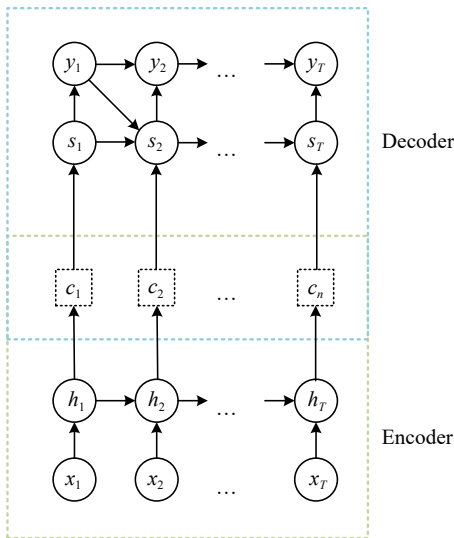


Fig. 7 Encoder-Decoder structure diagram.

avoids this problem by normalized initialization and intermediate normalization layers, to an extent that solves the deep network degradation problem. Not only does the ResNet network maintain the depth of the network, but it also prevents the degradation issue (Fig. 8). However, stacking ResNet blocks results in the problem of gradient disappearance or explosion, which affects model training speed and effectiveness^[67,68].

Multi-task learning

Multi-task learning is a joint learning method that identifies and appropriately measures relationships among tasks. This enables different tasks to provide each other with additional useful information for training models that perform better and are more robust. Multi-task learning relies on the common parameter among various tasks and the discovery of hidden common latent features between different tasks (Fig. 9). Conflicts or competition among different tasks may occur, resulting in reduced performance. Therefore, weighing task importance and considering the optimization goals of different tasks is necessary^[69,70].

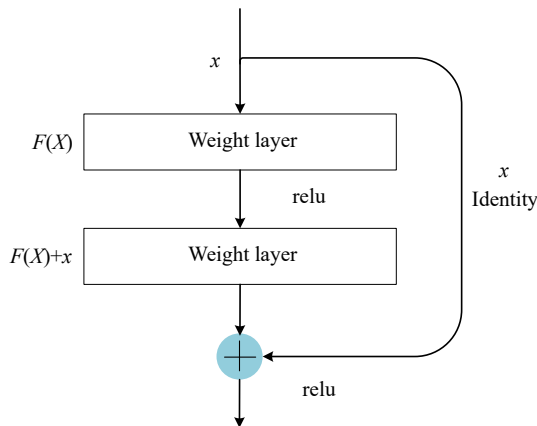


Fig. 8 Structure of ResNet network.

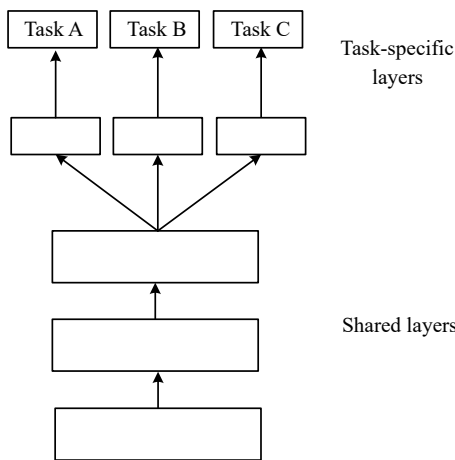


Fig. 9 Structure of multi-task learning.

Challenges

Challenge 1: Representation of dynamic correlations in OD flow

The paired attraction relationship between two regions is subject to dynamic changes over time, typically exhibiting stronger intensity during peak periods and weaker intensity during non-peak periods. Static graphs cannot represent the dynamic trend of OD flow. Therefore, capturing these relationships dynamically is crucial for node representation (Table 1).

In early taxi OD demand prediction studies, most research was based on static networks. Liu et al.^[72] constructed local spatial context (LSC) and global correlation context (GCC) modules based on Euclidean spatial grid data. The former learned the local spatial dependence of order demand from the starting point and destination perspectives, while the latter modeled the correlation between different regions. Wang et al.^[75] used grid embedding based on grid data to construct geographic and semantic neighbors to model passenger spatial flow patterns and adjacent relationships of different regions. The former measured the intrinsic closeness between grids and their neighbors, while the latter modeled the semantic intensity of traffic flow between starting points and destinations in the grid network. Chen et al.^[79] constructed the OD demands between each region in a single period based on regional grids. Then, they reduced the three-dimensional

tensor to a two-dimensional matrix through matrix cascading, considering the spatiotemporal properties in chronological order. Ke et al.^[80] encoded the context-aware spatial dependence of OD pairs by designing a residual multi-graph convolutional (RMGC) network through multiple OD graphs. Each node in the graph corresponded to an OD pair, and the adjacent matrix of the node was established to represent the neighborhood, distance, functional similarity, and historical demand correlation of OD pairs. The above studies represent the dependency relationship between regions based on static networks, ignoring the dynamic dependency relationships that may change over time.

Shi et al.^[78] constructed both static and dynamic graphs simultaneously to capture complex dynamic spatial dependency relationships and used the average strategy to obtain the final OD flow prediction. Zhang et al.^[81] proposed a dynamic node topology representation method to jointly represent the static and dynamic structural information of OD graphs. They introduced the k-TNEAT layer to adaptively adjust the relationship between each OD pair at different time intervals to learn the representation of nodes and edges, thus capturing the dynamic demand patterns of the time-varying OD graph. This method applies to both Euclidean and non-Euclidean datasets. Han et al.^[83] constructed a continuous time dynamic graph representation learning framework based on event updates, maintaining a dynamic state vector for each transportation node and representing multi-level spatiotemporal dependency relationships by sharing information among virtual cluster-level and regional-level nodes. Zhang et al.^[84] constructed dynamic graphs by treating the starting point and endpoint as two different semantic entities based on time updates, proposing an embedding module for the departure-destination pair and aggregating neighbor information through random walks. The above studies advance from predicting the starting point and destination on a static network to capturing spatiotemporal dynamic correlations by constructing dynamic graphs. Huang et al.^[91] developed a TMGCN layer to capture spatiotemporal correlations in dynamic OD graphs, which includes a static neighborhood relationship graph, Origin-Destination mutual attraction dynamic graph, and passengers' mobility association mode dynamic graph. This layer can learn relationships across different time intervals in all types of OD graphs.

Challenge 2: Spatial-temporal correlation

Spatial-temporal data possess both correlation and heterogeneity. Spatial-temporal correlation is manifested in the fact that each node can influence adjacent nodes at the next time step. Spatial-temporal heterogeneity is manifested in the different distributions of OD flow under conditions such as morning peak, evening peak, city center, and city edge. Currently, using two independent components to capture the temporal and spatial dependencies in a chained prediction often fails to capture the impact of spatial-temporal correlation and heterogeneity.

Zhang et al.^[84] and Han et al.^[83] applied node embedding on dynamic graphs, extending time into the spatial domain as the second dimension, which can simultaneously capture the structural relationship between nodes and their evolutionary relationship over time. Huang et al.^[87] proposed an OD attention mechanism to capture the unique spatial dependency between OD pairs with identical origins or destinations.

Review on taxi origin and destination prediction

Table 1. Summary of deep learning models in taxi origin-destination prediction.

Model	Spatial topology construction	Spatial dependency	Temporal dependency	Data set	Other factors
CSTN ^[71]	Raster	3DCNN	Conv LSTM	NYC-TOD	Local spatial context, meteorological information, globally relevant context
MultiConvLSTM ^[72]	Raster	MultiConv	ConvLSTM	NYC taxi	None
CLTS ^[73]	Raster	Conv2D	ConvLSTM	Beijing Taxi	None
GEML ^[74]	Raster (Geographic/semantic nodes)	SGCN (Grid embedding)	LSTM	NYC-taxi /DiDi ChengDu	Multi-task learning
FL-GCN ^[75]	Graph	Graph convolution (nodes, edges)	Kalman filtering	New Jersey Highway	None
CAS-CNN ^[76]	Raster	Split CNN		URT	Channel-wise attention
MPGCN ^[77]	Graph	2D-GCN	LSTM	DIDI Beijing /DiDi shanghai	None
GCN-SBULSTM ^[78]	Graph	GCN	Stacking bidirectional unidirectional LSTMs	SZ Metro	None
ST-ED-RMGC ^[79]	Graph	Multi-graph convolutional networks	LSTM	NYC taxi	Encoder decoder
DNEAT ^[80]	Dynamic node topology	GCN (k-TNEAT)	k-hop temporal encoder	DiDi ChengDu/ NYC taxi	None
Spatial OD-BiConvLSTM ^[81]	Raster	Conv2D	BiLSTM	NYC taxi	None
CMOD ^[82]	Graph (Event)	The graph represents learning	CTDG (Continuous-time evolution representation)	BJ Subway/NYC-Taxi	Multi- Head Attention
HMOD ^[83]	Graph (Event)	Graph embedded /Random walk	GRU/CTDG	NYC-Taxi/ Beijing Metro	None
SIZINB-GNN ^[84]	Graph	GNN	TCN	CDP dataset	None
ODformer ^[86]	Graph (Event)	2DGCN	ODformer	NYC taxi	OD attention
SI-GCN ^[87]	Graph (Event)	GCN (graph embedding)	Encoder-decoder	DIDI Beijing	a mapping function
STGDL ^[88]	Graph (road)	S-GCN	ResNet-based block	NYC taxi/DIDI Haikou	both short-term and long-term OD predictions
CWGAN-div ^[89]	Graph (road network)	GAN	ST-Conv CNN	NYC taxi	network-wide OD demand
DMGC-GAN ^[90]	Graph (neighbor/mutual attraction/passengers' mobility association mode)	GCN	ResNet	NYC taxi	None
Hex D-GCN ^[91]	Graph (hexagon-based path)	GCN	TMGCN	Taxi Shanghai	None
OD-TGAT ^[65]	Graph (grid map)	GAT	CNN	NYC Taxi	None
TFF ^[92]	Graph	GCN	GRU	Chongqing	A modified Kalman filter (KF)
CSGCN ^[93]	Graph	GCN	ST-Attention block	Taxi Beijing	Shifted Graph Clustering
gHMC-STA ^[94]	Graph	GCN	CNN	Taxi Beijing	Graph multi-head convolution for spatio-temporal aggregation
HSTN ^[95]	Raster	Separable 2D-CNN	Multi-Head Convolution	Taxi Shanghai	None
CTBGCN ^[96]	Graph	2DGCN	ResNet	NYC Taxi	None
CT-GCN ^[97]	Graph	GCN	Conv-LSTM	DIDI Haikou	None
			ST block		None

Challenge 3: Differentiation of different semantics of origin and destination

In a complex and irregular transportation network, the passenger demand of different OD pairs can be geographically and semantically correlated and has both directed and bidirectional correlations. However, modeling the demand separately for the origin and destination to learn local features around each grid discards the flow relationship between OD pairs and has no practical application. If only the distance and flow information between any two grids are considered without distinguishing the origin and destination, the directedness of the OD flow is ignored, and the varying attraction relationship between the origin and destination at different times is neglected.

Liu et al.^[72] constructed an LSC module that used two convolutional neural networks to learn local spatial contextual information of taxi demand from origin and destination views. However, this model does not take into account the flow relationship between the two regions or different semantic information. Wang et al.^[75] considered the combination of different origin-destination pairs and the number of passenger demands for each origin to predict the number of taxi orders from one area to another in a given time period, but only considered the flow relationship between two regions and ignored directedness and the distinction between different semantics and bi-directionality. Shi et al.^[78] constructed a multi-perspective graph convolutional network and proposed bidirectional correlations for OD flows when the start points are the

same or similar and the endpoints are the same or similar. Zhang et al.^[81] defined a weighted bidirectional graph and learned dynamic demand patterns from both the demand generation and attraction aspects while incorporating dynamic and bidirectional structure characteristics of edges. Zhang et al.^[84] proposed an origin-destination embedding module, treating the origin and destination as different semantic entities and using the parity of sampling to obtain semantic entities with different starting and ending points to distinguish different semantic information. Chen et al.^[79] proposed a BiConvLSTM method that processes input data in both forward and reverse directions through two ConvLSTMs, while maintaining hidden layer states and memory unit states in both directions.

Challenge 4: Time window selection

Currently, the predominant approach to predicting OD flows continues to be the discrete dynamic graph method for node prediction, which aggregates historical transactions into demand snapshots. Each snapshot contains demand within a fixed time window, resulting in disconnected OD flows. Moreover, the temporal aspect of OD flows is a continuous feature, and processing it under a fixed time window is intuitive but lacks rigor. The choice of time granularity can lead to biased prediction accuracy, with selecting too small a granularity generating a large amount of noise, and selecting too large a granularity causing decreased perception of important information. Additionally, predicting based on a continuous-time dynamic graph involves maintaining a dynamic state vector for each traffic node, potentially resulting in a large number of OD pairs and posing challenges in updating and maintaining representations for the many continuous time nodes.

Earlier approaches for OD demand forecasting aggregated taxi OD demand into demand snapshots, with each snapshot containing the total demand within a fixed time window. Zhang et al.^[81] designed a spatiotemporal attention network with a k-hop temporal node-edge attention layer to capture time-evolving node topology in dynamic OD graphs and to use different time granularities to explore complex time patterns, yet still falls under the category of discrete dynamic graph methods. Zhang et al.^[84] designed a layered memory storage technique that integrates discrete and continuous-time information of OD demand, extending the learning of traffic node representations to a continuous-time dynamic graph view. Han et al.^[83] constructed a framework for learning continuous-time dynamic graph representations, maintaining a dynamic state vector for each traffic node to store historical transaction information and continuously update it, lifting prediction from discrete time slices to continuous-time dynamic graph prediction.

Challenge 5: Data sparseness problem solving

Each OD pair has a time sequence that requires more complex spatial dependencies. Discrete dynamic graph-based prediction methods inevitably suffer from information loss and produce a large number of zero values. The use of continuous-time dynamic graph-based prediction methods also encounters the problem of sparse data for some OD pairs, and exacerbates the data sparsity issue through the quadratic quantity of predicted OD demand.

Wang et al.^[75] proposed a pre-weighted aggregator that combines the perception of data sparsity and range at different granularity levels based on grid embedding to mitigate the

impact of sparse data. Hu et al.^[86] designed three modules, namely, factorization, prediction, and restoration, to handle data sparsity in matrix factorization and restoration steps. Zhang et al.^[77] introduced a segmentation CNN to transform sparse OD data into dense features and verified the effectiveness of a masking loss function for tackling data dimensionality and sparsity issues. Zhang et al.^[81] designed a loss function \downarrow_{reg} , commonly used for quadratic regression loss, and a masking loss function \downarrow_{mask} that prioritizes harder-to-predict edges based on the auxiliary loss function approach to combat the adverse effects of high sparsity. Zhuang et al.^[85] proposed a STZINB-GNN model to quantify the uncertainty of sparse travel demand using the zero-inflated negative binomial (ZINB) distribution to capture an extensive amount of zeros in sparse O-D matrices and the negative binomial (NB) distribution for each non-zero entry. The model also introduced a spatiotemporal embedding with an additional parameter π to learn the likelihood of input zeros. Han et al.^[83] mitigated sparse data issues by proposing a layered message passing module, allowing virtual cluster-level nodes and region-level nodes to share information, and designing a loss function that focuses more on non-zero demand. Yao et al.^[88] applied two gating mechanisms to the vanilla convolution operation to alleviate the error accumulation issue of typical recurrent forecasting in long-term OD prediction. Zou et al.^[89] proposed using residual blocks and refined loss functions to enhance model training stability. Huang et al.^[91] used a GAN structure to address the problem of zero-valued elements dominating the OD demand matrix. Yang et al.^[65] used a method based on filling the lower triangular matrix, only considering OD pairs with travel volumes greater than zero, and for the remaining OD pairs, used a method similar to interpolation. Li et al.^[93] Proposed a hybrid framework to predict short-term OD, in which the two-step design predicts trip generation/attraction in the first stage, and estimates trip distribution in the second stage. This framework not only takes advantage of robust spatio-temporal predictors but also avoids under- or over-estimations of short-term OD matrices due to high sparsity. These 5 challenges are reviewed and summarized in Table 2.

Public datasets and open-source codes

Details shown in Tables 3 and 4.

Summary and outlook

This review comprehensively examines the departure-arrival prediction problem using deep learning algorithms. Specifically, we summarize deep learning methods for taxi demand prediction from four aspects: topology construction, spatial dependency, temporal dependency, and other factors. In addition, based on the decomposition of the studied architecture, we summarize common challenges in departure-arrival prediction such as dynamics, spatiotemporal dependencies, semantic differentiation, time window selection, and data sparsity. Importantly, we provide multiple existing solutions for each challenge. Finally, we provide hyperlinks to public datasets and codes of related work to facilitate future research. We also propose future directions for those interested in this field.

Review on taxi origin and destination prediction

Table 2. Summary of problem solving in taxi origin-destination prediction.

Model	Dynamic/Static	Directed/Undirected	Time window	Sparse data	Spatial-temporal correlation
GEML ^[74]	Static	Fluid relationship but undirected	Discrete-time snapshots with the same time granularity	Multi-granularity level mesh embedding/pre-weighted aggregator	None
MultiConvLSTM ^[72]	Dynamic and Static	Undirected	Discrete-time snapshots with the same time granularity	Self-attention	None
CLTS ^[73]	Dynamic	Undirected	Discrete-time snapshots with the same time granularity	None	None
CSTN ^[71]	Static	Undirected	Discrete-time snapshots with the same time granularity	None	None
CAS-CNN ^[76]	Static	Undirected	Discrete-time snapshots with the same time granularity	Split the CNN/masking loss function	None
MPGCN ^[77]	Two static adjacency matrices and one dynamic adjacency	Undirected	Discrete-time snapshots with the same time granularity	None	Spatiotemporal dynamic adjacency matrix
GCN-SBULSTM ^[78]	Static	Undirected	Discrete-time snapshots with the same time granularity	None	None
ST-ED-RMGC ^[79]	Static	Undirected	Discrete-time snapshots with the same time granularity	None	Parallel prediction
DNEAT ^[80]	Dynamic and Static	Directed	Discrete-time snapshots of different time granularities	The loss function \downarrow jointly minimizes the \downarrow_{reg} and \downarrow_{mask} masks	None
Spatial OD-BiConvLSTM ^[81] CMOD ^[82]	Dynamic	Undirected	Discrete-time snapshots (sliding window)	The loss function \downarrow	None
	Dynamic	Undirected	Continuous dynamic time	The loss function focuses on non-zero demand	Node embedding
HMOD ^[83]	Dynamic and Static	Directed and semantically differentiated	Continuous dynamic time	None	Node embedding
SIZINB-GNN ^[84]	Static	Undirected	Discrete-time snapshots with the same time granularity	Zero expansion negative binomial distribution/probability of learning input being zero additive parameter π	None
ODformer ^[86]	Static	Directed	Long sequence time window	Transformer	Spatial-Temporal Transformers
SI-GCN ^[87]	Dynamic	Directed	Continuous dynamic time	Negative sampling	data imputation,
STGDL ^[88]	Static	Directed	Discrete-time snapshots (sliding window)	Two gate mechanisms	ST-GDL model
CWGAN-div ^[89]	Dynamic and Static	Undirected	Discrete-time snapshots (moving average)	Residual blocks	Interpretable conditional information
DMGC-GAN ^[90]	Dynamic and Static	Directed	Discrete-time snapshots (sliding window)	GAN	TMGCN
Hex D-GCN ^[91]	Dynamic	Undirected	Discrete-time snapshots (moving average)	Filling the lower triangle matrix	None
OD-TGAT ^[65]	Static	Directed	Discrete-time snapshots (sliding window)	GAT	None
TFF ^[92]	Dynamic	Directed	Discrete-time snapshots (moving average)	Two-step design performs	None
CSGCN ^[93]	Static	Directed	Discrete-time snapshots (sliding window)	None	None
gHMC-STA ^[94]	Static	Directed	Discrete-time snapshots (sliding window)	None	None
HSTN ^[95]	Static	Directed	Discrete-time snapshots (moving average)	None	None
CTBGCN ^[96]	Static	Directed	Discrete-time snapshots (sliding window)	None	None
CT-GCN ^[97]	Static	Directed	Discrete-time snapshots (sliding window)	None	None

Apply the relevant direction

Spatiotemporal dynamic correlation

In a dynamic road network, the spatiotemporal dependency of an individual node is affected by the overall interaction and

randomness of the network. At present, research mainly addresses the spatiotemporal dynamics issue in traffic flow prediction by introducing attention mechanisms. Therefore, further investigation into the application of attention

Table 3. Open taxi datasets.

Dataset	Links
NYC-Taxi	https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page
Porto	–
T-drive (Beijing)	https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/
Taxi-Shanghai	–
Taxi-Shenzhen	https://opendata.sz.gov.cn/
Taxi-Chengdu	https://tianchi.aliyun.com/dataset/39384

Table 4. Open-source codes.

Model	Github
GEML	https://github.com/Zekun-Cai/GEML-Origin-Destination-Matrix-Prediction-via-Graph-Convolution
CSTN	https://github.com/liulingbo918/CSTN
FL-GCN	https://github.com/alzmx/OD_Prediction
MPGCN	https://github.com/underdoc-wang/MPGCN
ST-ED-RMGC	https://github.com/kejintao/ST-ED-RMGC/tree/main/od_prediction
CMOD	https://github.com/liangzhehan/cmod
HMOD	https://github.com/Rising0321/HMOD
SIZINB-GNN	https://github.com/zhuangdingyi/stzinb

mechanisms in predicting taxi OD demand models that combine spatiotemporal features could be explored more deeply.

External information addition

External factors, such as holidays, weather, points of interest (POI), large events, and traffic accidents, also have a significant impact on taxi demand prediction. However, many existing OD demand prediction models rarely consider external factors, which are diverse and difficult to collect, and suffer from sparsity issues. Therefore, how to effectively handle external factors and maximize their contribution to the prediction remains a challenge in the research community.

Regional division

So far, both grid-based and graph-based OD demand prediction methods rely on manually selected spatial data, whether it is dividing the area into grids or traffic zones. This approach is intuitive and convenient but lacks rigor, and it is impossible to list all potential relationships by human design, which limits the generalization ability of the model. Therefore, there is still a need for extensive research on how to partition regions in a reasonable and non-manual way when facing a completely new area.

OD data sparseness and data overload

Compared to regional demand prediction, the abundance of zero values and sparsity are still significant challenges in OD data prediction. Additionally, continuous-time dynamic prediction also leads to second-order stations, highlighting the need for further exploration on how to handle them during model computation and allow available data to be maximally utilized.

Technical related issues

Most existing approaches to prediction tasks utilize recurrent neural networks (RNNs) and graph convolutional networks (GCNs), with only a few studies employing graph attention networks (GATs) or graph autoencoders (GAEs). Therefore, further research is needed to investigate how other advanced

graph neural network models can be applied to OD demand prediction problems and expanded upon to better suit traffic prediction.

Acknowledgments

This work was supported by 2022 Shenyang Philosophy and Social Science Planning under grant SY202201Z, Liaoning Provincial Department of Education Project under grant LJKZ0588.

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 21 April 2023; Accepted 22 August 2023; Published online 28 September 2023

References

1. Tebaldi C, West M. 1998. Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association* 93:557–73
2. Carvalho L. 2014. A Bayesian statistical approach for inference on static origin–destination matrices in transportation studies. *Techonometrics* 56:225–37
3. Spiess H. 1987. A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B: Methodological* 21:395–412
4. Chang GL, Tao X. 1999. An integrated model for estimating time-varying network origin-destination distributions. *Transportation Research Part A: Policy and Practice* 33:381–99
5. Chen Y, Ordóñez F, Palmer K. 2006. Confidence intervals for OD demand estimation. *USC-ISE Working Paper* 2006:1
6. Hazelton ML. 2008. Statistical inference for time varying origin-destination matrices. *Transportation Research Part B: Methodological* 42:542–52
7. Djukic T, Flötteröd G, van Lint H, Hoogendoorn S. 2012. Efficient real time OD matrix estimation based on Principal Component Analysis. 2012 15th International IEEE Conference on Intelligent Transportation Systems, Anchorage, AK, USA, 2012. USA: IEEE. pp. 115–21. <https://doi.org/10.1109/ITSC.2012.6338720>
8. Shao H, Lam WHK, Sumalee A, Chen A, Hazelton ML. 2014. Estimation of mean and covariance of peak hour origin-destination demands from day-to-day traffic counts. *Transportation Research Part B: Methodological* 68:52–75
9. Lu S, Wang J, Xue Z, Liu X. 2016. Traffic analysis and OD travel time matrix based on two-fluid model. *Journal of Highway and Transportation Research and Development (English Edition)* 10:78–84
10. Zhu X, Guo D. 2017. Urban event detection with big data of taxi OD trips: a time series decomposition approach. *Transactions in GIS* 21:560–74
11. Ren J, Xie Q. 2017. Efficient OD trip matrix prediction based on tensor decomposition. 2017 18th IEEE International Conference on Mobile Data Management (MDM), Daejeon, Korea (South), 2017. UAS: IEEE. pp. 180–85. <https://doi.org/10.1109/MDM.2017.32>
12. Li X, Kurths J, Gao C, Zhang J, Wang Z, et al. 2017. A hybrid algorithm for estimating origin-destination flows. *IEEE Access* 6:677–87
13. Li J, Wen H, Lin L, Qi W. 2018. Demand prediction model of E-hailing based on QPSO_RBF neural network. *Journal of Guangxi University (Natural Science Edition)* 43(2):700–9
14. Lu Y, Li S. 2014. An empirical study of with-in day OD prediction using taxi GPS data in Singapore. Report. No. 14-5074.

Review on taxi origin and destination prediction

15. Hong WC. 2011. Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm. *Neurocomputing* 74(12–13):2096–107
16. Tong Y, Chen Y, Zhou Z, Chen L, Wang J, et al. 2017. The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms. *KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 2017*. New York, United States: Association for Computing Machinery. pp. 1653–62. <https://doi.org/10.1145/3097983.3098018>
17. Skarding J, Gabrys B, Musial K. 2021. Foundations and modeling of dynamic networks using dynamic graph neural networks: a survey. *IEEE Access* 9:79143–68
18. Huang H, Fang Z, Wang X, Miao Y, Jin H. 2020. Motif-Preserving Temporal Network Embedding. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, Yokohama, Japan, 2020*. California: International Joint Conferences on Artificial Intelligence Organization. pp. 1237–43. <https://doi.org/10.24963/ijcai.2020/172>
19. Trivedi R, Farajtabar M, Biswal P, et al. 2019. Dyrep: Learning representations over dynamic graphs. *International Conference on Learning Representations*.
20. Kumar S, Zhang X, Leskovec J. 2019. Predicting dynamic embedding trajectory in temporal interaction networks. *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 2019*. New York, United States: Association for Computing Machinery. pp. 1269–78. <https://doi.org/10.1145/3292500.3330895>
21. Lv Y, Duan Y, Kang W, Li Z, Wang FY. 2015. Traffic flow prediction with big data: a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems* 16:865–73
22. Krupski J, Graniszewski W, Iwanowski M. 2021. Data transformation schemes for CNN-based network traffic analysis: a survey. *Electronics* 10:2042
23. Ranjan N, Bhandari S, Zhao HP, Kim H, Khan P. 2020. City-wide traffic congestion prediction based on CNN, LSTM and transpose CNN. *IEEE Access* 8:81606–20
24. Li X, Zhao Z, Wang Q. 2022. ABSSNet: attention-based spatial segmentation network for traffic scene understanding. *IEEE Transactions on Cybernetics* 52:9352–62
25. Baheti B, Gajre S, Talbar S. 2019. Semantic scene understanding in unstructured environment with deep convolutional neural network. *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), 2019, Kochi, India, 2019*. USA: IEEE. pp. 790–95. <https://doi.org/10.1109/TENCON.2019.8929376>
26. Haque WA, Arefin S, Shihavuddin ASM, Hasan MA. 2021. DeepThin: a novel lightweight CNN architecture for traffic sign recognition without GPU requirements. *Expert Systems with Applications* 168:114481
27. Zhang J, Xie Z, Sun J, Zou X, Wang J. 2020. A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection. *IEEE Access* 8:29742–54
28. Bogaerts T, Masegosa AD, Angarita-Zapata JS, Onieva E, Hellinckx P. 2020. A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. *Transportation Research Part C: Emerging Technologies* 112:62–77
29. Zhou Z, Qin Y, Luo H. 2021. Deep spatio-temporal convolutional neural network for city traffic flow prediction. *2021 2nd International Conference on Computing and Data Science (CDS), Stanford, CA, USA, 2021*. USA: IEEE. pp. 171–75. <https://doi.org/10.1109/CDS52072.2021.00037>
30. Guo S, Lin Y, Li S, Chen Z, Wan H. 2019. Deep spatial-temporal 3D convolutional neural networks for traffic data forecasting. *IEEE Transactions on Intelligent Transportation Systems* 20:3913–26
31. Ma X, Dai Z, He Z, Ma J, Wang Y, et al. 2017. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors* 17:818
32. Ran J, Chen Y, Li S. 2019. Three-dimensional convolutional neural network based traffic classification for wireless communications. *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Anaheim, CA, USA, 2018*. USA: IEEE. pp. 624–27. <https://doi.org/10.1109/GlobalSIP.2018.8646659>
33. Zhu J, Wang Q, Tao C, Deng H, Zhao L, et al. 2021. AST-GCN: attribute-augmented spatiotemporal graph convolutional network for traffic forecasting. *IEEE Access* 9:35973–83
34. Li Z, Xiong G, Chen Y, Lv Y, Hu B, et al. 2019. A hybrid deep learning approach with GCN and LSTM for traffic flow prediction. *2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 2019*. USA: IEEE. pp. 1929–33. <https://doi.org/10.1109/ITSC.2019.8916778>
35. Diao Z, Xie G, Wang X, Ren R, Meng X, et al. 2023. EC-GCN: a encrypted traffic classification framework based on multi-scale graph convolution networks. *Computer Networks* 224:109614
36. Guo K, Hu Y, Sun Y, Qian S, Gao J, et al. 2021. Hierarchical graph convolution network for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 35:151–59
37. Dong X, Thanou D, Rabbat M, Frossard P. 2019. Learning graphs from data: a signal representation perspective. *IEEE Signal Processing Magazine* 36:44–63
38. Geng X, Li Y, Wang L, Zhang L, Yang Q, et al. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 33:3656–63
39. Cui Z, Henrickson K, Ke R, Wang Y. 2020. Traffic graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems* 21:4883–94
40. Ali A, Zhu Y, Chen Q, Yu J, Cai H. 2020. Leveraging spatio-temporal patterns for predicting citywide traffic crowd flows using deep hybrid neural networks. *2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS), Tianjin, China, 2019*. USA: IEEE. pp. 125–32. <https://doi.org/10.1109/ICPADS47876.2019.00025>
41. Yu L, Du B, Hu X, Sun L, Han L, et al. 2021. Deep spatio-temporal graph convolutional network for traffic accident prediction. *Neurocomputing* 423:135–47
42. Li M, Zhu Z. 2021. Spatial-temporal fusion graph neural networks for traffic flow forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 35:4189–96
43. Wang X, Ma Y, Wang Y, Jin W, Wang X, et al. 2020. Traffic flow prediction via spatial temporal graph neural network. *WWW '20: Proceedings of The Web Conference 2020, Taipei, Taiwan, 2020*. New York, United States: Association for Computing Machinery. pp. 1082–92. <https://doi.org/10.1145/3366423.3380186>
44. Zhang Q, Yu K, Guo Z, Garg S, Rodrigues JJPC, et al. 2021. Graph neural network-driven traffic forecasting for the connected internet of vehicles. *IEEE Transactions on Network Science and Engineering* 9(5):3015–27
45. Liu T, Wu W, Zhu Y, Tong W. 2020. Predicting taxi demands via an attention-based convolutional recurrent neural network. *Knowledge-Based Systems* 206:106294
46. Rossi A, Barlacchi G, Bianchini M, Lepri B. 2020. Modelling taxi drivers' behaviour for the next destination prediction. *IEEE Transactions on Intelligent Transportation Systems* 21:2980–89
47. Tian Y, Pan L. 2016. Predicting short-term traffic flow by long short-term memory recurrent neural network. *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), Chengdu, China 2015*. USA: IEEE. pp. 153–58. <https://doi.org/10.1109/SmartCity.2015.63>
48. Fukuda S, Uchida H, Fujii H, Yamada T. 2020. Short-term prediction of traffic flow under incident conditions using graph convolutional recurrent neural network and traffic simulation. *IET Intelligent Transport Systems* 14:936–46

49. Kim K, Lee JH, Lim HK, Oh S, Han YH. 2022. Deep RNN-based network traffic classification scheme in edge computing system. *Computer Science and Information Systems* 19:165–84
50. Paul A, Mitra S. 2021. Management of traffic signals using deep reinforcement learning in bidirectional recurrent neural network in ITS. *ISMSI '21: Proceedings of the 2021 5th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence, Victoria, Seychelles, 2021*. New York, United States: Association for Computing Machinery. pp. 60–64. <https://doi.org/10.1145/3461598.3461608>
51. Li M, Wang Y, Wang Z, Zheng H. 2020. A deep learning method based on an attention mechanism for wireless network traffic prediction. *Ad Hoc Networks* 107:102258
52. Lai Y, Zhang K, Lin J, Yang F, Fan Y. 2020. Taxi demand prediction with LSTM-based combination model. *2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCLOUD/SocialCom/SustainCom), Xiamen, China, 2019*. USA: IEEE. pp. 944–50. <https://doi.org/10.1109/ISPA-BDCLOUD-SustainCom-SocialCom48970.2019.00137>
53. Nihale S, Sharma S, Parashar L, Singh U. 2020. Network traffic prediction using long short-term memory. *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020*. USA: IEEE. pp. 338–43. <https://doi.org/10.1109/ICESC48915.2020.9156045>
54. Zeng C, Ma C, Wang K, Cui Z. 2022. Predicting vacant parking space availability: a DWT-Bi-LSTM model. *Physica A: Statistical Mechanics and Its Applications* 599:127498
55. Fu R, Zhang Z, Li L. 2017. Using LSTM and GRU neural network methods for traffic flow prediction. *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China, 2016*. USA: IEEE. pp. 324–28. <https://doi.org/10.1109/YAC.2016.7804912>
56. Zhao J, Kong W, Zhou M, Zhou T, Xu Y, et al. 2022. Prediction of urban taxi travel demand by using hybrid dynamic graph convolutional network model. *Sensors* 22:5982
57. Abideen ZU, Sun H, Yang Z, Ahmad RZ, Iftekhhar A, et al. 2020. Deep wide spatial-temporal based transformer networks modeling for the next destination according to the taxi driver behavior prediction. *Applied Sciences* 11:17
58. Tsiligkaridis A, Zhang J, Taguchi H, Nikovski D. 2020. Personalized destination prediction using transformers in a contextless data setting. *2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020*. USA: IEEE. pp. 1–7. <https://doi.org/10.1109/IJCNN48605.2020.9207514>
59. Li D, Lin C, Gao W, Chen Z, Wang Z, et al. 2020. Capsules TCN network for urban computing and intelligence in urban traffic prediction. *Wireless Communications and Mobile Computing* 2020:6896579
60. Wang Y, Li J, Zhao A, Lv Z, Lu G. 2021. Temporal attention-based graph convolution network for taxi demand prediction in functional areas. *WASA 2021: Wireless Algorithms, Systems, and Applications, Nanjing, China, 2021*. Switzerland: Springer, Cham. pp. 203–14. https://doi.org/10.1007/978-3-030-85928-2_16
61. Xu J, Rahmatizadeh R, Bölöni L, Turgut D. 2018. Real-time prediction of taxi demand using recurrent neural networks. *IEEE Transactions on Intelligent Transportation Systems* 19:2572–81
62. Chang HW, Tai YC, Hsu JYJ. 2010. Context-aware taxi demand hotspots prediction. *International Journal of Business Intelligence and Data Mining* 5:3–18
63. Tong Y, Chen Y, Zhou Z, Chen L, Wang J, et al. 2017. The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms. *KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 2017*. New York, United States: Association for Computing Machinery. pp. 1653–62. <https://doi.org/10.1145/3097983.3098018>
64. Vanichrujee U, Horanont T, Pattara-atikom W, Theeramunkong T, Shinozaki T. 2018. Taxi demand prediction using ensemble model based on RNNs and XGBOOST. *2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES), Khon Kaen, Thailand, USA: IEEE*. pp. 1–6. <https://doi.org/10.1109/ICESIT-ICICTES.2018.8442063>
65. Xu Y, Li D. 2019. Incorporating graph attention and recurrent architectures for city-wide taxi demand prediction. *ISPRS International Journal of Geo-Information* 8:414
66. Liu Y, Liu Z, Lyu C, Ye J. 2020. Attention-based deep ensemble net for large-scale online taxi-hailing demand prediction. *IEEE Transactions on Intelligent Transportation Systems* 21:4798–807
67. Kuang L, Yan X, Tan X, Li S, Yang X. 2019. Predicting taxi demand based on 3D convolutional neural network and multi-task learning. *Remote Sensing* 11:1265
68. Duan ZT, Zhang K, Yang Y, Ni YY, Saurab B. 2018. Taxi demand prediction based on CNN-LSTM-ResNet hybrid depth learning model. *Journal of Transportation Systems Engineering and Information Technology* 18(4):215–23
69. Zhang C, Zhu F, Wang X, Sun L, Tang H, et al. 2022. Taxi demand prediction using parallel multi-task learning model. *IEEE Transactions on Intelligent Transportation Systems* 23:794–803
70. Chen Z, Zhao B, Wang Y, Duan Z, Zhao X. 2020. Multitask learning and GCN-based taxi demand prediction for a traffic road network. *Sensors* 20:3776
71. Liu L, Qiu Z, Li G, Wang Q, Ouyang W, et al. 2019. Contextualized spatial-temporal network for taxi origin-destination demand prediction. *IEEE Transactions on Intelligent Transportation Systems* 20:3875–87
72. Duan Z, Zhang K, Chen Z, Liu Z, Tang L, et al. 2019. Prediction of city-scale dynamic taxi origin-destination flows using a hybrid deep neural network combined with travel time. *IEEE Access* 7:127816–32
73. Chu KF, Lam AYS, Li VOK. 2020. Deep multi-scale convolutional LSTM network for travel demand and origin-destination predictions. *IEEE Transactions on Intelligent Transportation Systems* 21:3219–32
74. Wang Y, Yin H, Chen H, Wo T, Xu J, et al. 2019. Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 4 - 8, 2019, Anchorage, AK, USA*. ACM: 1227–35
75. Xiong X, Ozbay K, Jin L, Feng C. 2020. Dynamic origin-destination matrix prediction with line graph neural networks and Kalman filter. *Transportation Research Record: Journal of the Transportation Research Board* 2674:491–503
76. Zhang J, Che H, Chen F, Ma W, He Z. 2020. Short-term origin-destination demand prediction in urban rail transit systems: a channel-wise attentive split-convolutional neural network method. *arXiv* In press
77. Shi H, Yao Q, Guo Q, Li Y, Zhang L, et al. 2020. Predicting origin-destination flow via multi-perspective graph convolutional network. *2020 IEEE 36th International Conference on Data Engineering (ICDE), Dallas, TX, USA, 2020*. USA: IEEE. pp. 1818–21. <https://doi.org/10.1109/ICDE48307.2020.00178>
78. Chen P, Fu X, Wang X. 2022. A graph convolutional stacked bidirectional unidirectional-LSTM neural network for metro ridership prediction. *IEEE Transactions on Intelligent Transportation Systems* 23:6950–62
79. Ke J, Qin X, Yang H, Zheng Z, Zhu Z, et al. 2021. Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network. *Transportation Research Part C: Emerging Technologies* 122:102858
80. Zhang D, Xiao F, Shen M, Zhong S. 2021. DNEAT: a novel dynamic node-edge attention network for origin-destination demand

Review on taxi origin and destination prediction

- prediction. *Transportation Research Part C: Emerging Technologies* 122:102851
81. Chen D, Wang J, Xiong C. 2021. Research on origin-destination travel demand prediction method of inter-regional online taxi based on SpatialOD-BiConvLSTM. *IET Intelligent Transport Systems* 15:1533–47
 82. Han L, Ma X, Sun L, Du B, Fu Y, et al. 2022. Continuous-time and multi-level graph representation learning for origin-destination demand prediction. KDD '22: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington DC, USA, 2022. New York, United States: Association for Computing Machinery. pp. 516–24. <https://doi.org/10.1145/3534678.3539273>
 83. Zhang R, Han L, Liu B, Zeng J, Sun L. 2022. Dynamic graph learning based on hierarchical memory for origin-destination demand prediction. *arXiv* In press
 84. Zhuang D, Wang S, Koutsopoulos H N, et al. 2022. Uncertainty quantification of sparse travel demand prediction with spatial-temporal graph neural networks. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*, Washington DC, USA, 2022. New York, United States: Association for Computing Machinery. pp. 4639–47. <https://doi.org/10.1145/3534678.3539093>
 85. Hu J, Yang B, Guo C, Jensen CS, Xiong H. 2020. Stochastic origin-destination matrix forecasting using dual-stage graph convolutional, recurrent neural networks. *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, Dallas, TX, USA, 2020. USA: IEEE. pp. 1417–28. <https://doi.org/10.1109/ICDE48307.2020.00126>
 86. Huang B, Ruan K, Yu W, Xiao J, Xie R, et al. 2023. ODformer: spatial-temporal transformers for long sequence Origin-Destination matrix forecasting against cross application scenario. *Expert Systems with Applications* 222:119835
 87. Yao X, Gao Y, Zhu D, Manley E, Wang J, et al. 2021. Spatial origin-destination flow imputation using graph convolutional networks. *IEEE Transactions on Intelligent Transportation Systems* 22:7474–84
 88. Zou X, Zhang S, Zhang C, Yu JJQ, Chung E. 2022. Long-term origin-destination demand prediction with graph deep learning. *IEEE Transactions on Big Data* 8:1481–95
 89. Wang N, Zheng L, Shen H, Li S. 2023. Ride-hailing origin-destination demand prediction with spatiotemporal information fusion. *Transportation Safety and Environment* Accepted paper:tdad026
 90. Huang Z, Zhang W, Wang D, Yin Y. 2022. A GAN framework-based dynamic multi-graph convolutional network for origin-destination-based ride-hailing demand prediction. *Information Sciences* 601:129–46
 91. Yang Y, Zhang S, Zhang C, Yu JJQ. 2021. Origin-destination matrix prediction via hexagon-based generated graph. *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, Indianapolis, IN, USA, 2021. USA: IEEE. pp. 1399–404. <https://doi.org/10.1109/ITSC48978.2021.9564718>
 92. Li D, Wang W, Zhao D. 2023. Designing a novel two-stage fusion framework to predict short-term origin-destination flow. *Journal of Transportation Engineering-Part A: Systems* 149(5):04023032
 93. Peng Z, Wu G, Xia F. 2021. Clustering shift graph convolutional network for taxi origin-destination demand prediction. *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, Washington, DC, USA, 2021. USA: IEEE. pp. 268–72. <https://doi.org/10.1109/ICTAI52525.2021.00044>
 94. Bhanu M, Kumar R, Roy S, Mendes-Moreira J, Chandra J. 2022. Graph multi-head convolution for spatio-temporal attention in origin destination tensor prediction. In *PAKDD 2022: Advances in Knowledge Discovery and Data Mining*, eds. Gama J, Li T, Yu Y, Chen E, Zheng Y, et al. Switzerland: Springer Cham. pp. 459–71. https://doi.org/10.1007/978-3-031-05933-9_36
 95. Chen T, Nie L, Pan J, Tu L, Zheng B, et al. 2023. Origin-destination traffic prediction based on hybrid spatio-temporal network. *2022 IEEE International Conference on Data Mining (ICDM)*, Orlando, FL, USA, 2022. USA: IEEE. pp. 879–84. <https://doi.org/10.1109/ICDM54844.2022.00101>
 96. Cao Y, Liu L, Dong Y. 2023. Convolutional long short-term memory two-dimensional bidirectional graph convolutional network for taxi demand prediction. *Sustainability* 15:7903
 97. Shuai C, Zhang X, Wang Y, He M, Yang F, et al. 2023. Online car-hailing origin-destination forecast based on a temporal graph convolutional network. *IEEE Intelligent Transportation Systems Magazine* 15:121–36



Copyright: © 2023 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.