





Research Article

Knowledge Graph-Based Enhanced Transformer for Metro Individual Travel Destination Prediction

Hainan Chi ^{1,2}, Boyue Wang ^{1,2}, Qibin Ge ³, and Guangyu Huo ^{1,2}

¹Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing 100124, China

²Beijing Artificial Intelligence Institute, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

³Beijing Intelligent Transportation Development Center, Beijing 102208, China

Correspondence should be addressed to Boyue Wang; wby@bjut.edu.cn

Received 4 January 2022; Accepted 28 February 2022; Published 4 April 2022

Academic Editor: Yanming Shen

Copyright © 2022 Hainan Chi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate and timely destination prediction of subway passengers is of great significance in improving urban residents' travel efficiency, alleviating urban traffic pressure, and recommending the proper location-based service. Although some individual travel destination prediction methods have been proposed, the prediction performance is poor due to the large difference in travel locations of different individuals, the difficulty of evaluating the individual travel intention, the sparsity of individual travel trajectory data, and other problems. To solve these problems, this paper proposes a knowledge graph-based enhanced Transformer method (KG-Trans) for the metro individual travel destination prediction task (MITD-Pre), which contains three main modules: (1) the knowledge graph (KG) module constructs a multilayer individual travel KG from top to bottom, which accurately describes the travel individuals and their travel intentions. By analyzing the association relationship between nodes in the KG, the relationship between travel individuals can be naturally established. The learned similar travel regularity can solve the problem of sparse travel trajectories of some individuals. (2) The enhanced Transformer module extracts the dynamic and hierarchical features from the long-term sequential travel trajectory data. (3) The classifier module introduces the cross-entropy loss to constrain the uniqueness of the predicted subway travel station. The experimental results show that the proposed method obtains a higher destination prediction accuracy than the previous individual travel destination prediction methods.

1. Introduction

As an important part of public transport, the urban rail transit produces a large amount of spatiotemporal trajectory data in real time, which contains rich spatiotemporal location information and reflects the travel mode of passengers. This gives us an opportunity to deeply explore the individual travel patterns and regularity. Traffic prediction is a very basic and important problem in the field of transportation. Most existing traffic prediction methods focus on traffic flow, speed, and so on [1–3]. However, with the development of information technology and various intelligent devices, strong data support and technical support are created for individual travel destination prediction. The real-time prediction of the travel destination of each individual

who stays in the subway station is of great significance for the tracking of individuals, service recommendations, and the construction of the smart city. And, it is bound to become an important social demand in the era of big data.

At present, a few methods have been proposed in the field of travel destination prediction, such as the Markov model [4], Bayesian model [5], and Gaussian mixture model [6]. These methods predict the travel destination according to the general mobility characteristics of individuals. However, they ignore the differences in individual behaviors between users [7, 8] and the problems of individual travel data, such as the high spatial complexity and sparse historical travel trajectory. Therefore, their prediction results are unsatisfactory. In addition, considering that there is still a huge challenge in the task of individual travel destination

prediction, that is, how to accurately grasp the individual travel intention, travel intention may be affected by time, location, and other factors. For example, when individuals travel to the same place, their travel intentions may be different on weekdays and weekends. These challenges are beyond the previous methods.

As we all know, KG is a very advanced carrier containing a lot of common-sense knowledge and plays an important role in many practical applications. The emergence of KG provides a new perspective to comprehensively describe individual travel patterns. It carries out much application research; e.g., the entity portrait and the law prediction tasks are achieved by utilizing relationship reasoning and knowledge aggregation. So, KG provides a new method of support for accurately quantifying individual travel patterns in the public transport. It effectively breaks through the traditional expression limitations based on the traffic big data.

Inspired by the KG, this paper integrates deep learning prediction and the KG to achieve accurate travel destination prediction tasks. Specifically, we construct an individual travel KG based on the historical travel data of individuals and then conduct the portrait analysis based on the KG to accurately grasp the individual travel intention. At the same time, for the historical travel trajectory data with the long-term time series and the long-term time dependence, Transformer is used to learn the dynamic and hierarchical characteristics in the sequence data to achieve the final prediction tasks.

This study mainly aims to integrate the KG into the individual travel destination prediction model. The main contributions of this study are summarized as follows:

- (i) An individual travel KG is constructed, and we propose a novel individual travel destination prediction method based on such KG, which aims to accurately analyze the individual travel patterns and intentions
- (ii) We analyze the travel groups having similar travel trajectories to handle the sparse historical trajectories of some individuals, which is obviously different from the traditional individual travel destination prediction methods
- (iii) An enhanced Transformer module is proposed to extract the dynamic and hierarchical features in the history of travel trajectory data
- (iv) Experimental results show that the proposed method effectively exploits KG to analyze the subway card data and obtains satisfactory performance

2. Related Work

In this section, we review several related types of research about the KGs and the individual travel destination predictions.

2.1. Knowledge Graph. KG is a very advanced carrier having a lot of common-sense knowledge. And, it acquires success in many practical applications, such as knowledge questions

and answers and medical fields. Several general KGs established by Baidu, Google, or other organizations also play an important role in our daily lives. At present, there are many semantic knowledge bases based on Wikipedia in the construction of general domain KG abroad, such as Freebase [9], DBpedia [10], and Yago. The general KGs in China are the bilingual encyclopedic KG XLORE [11] developed by Tsinghua University, the Chinese general encyclopedia KG CN-DBpedia developed by Fudan University, and Zhishi.me [12] developed by Shanghai Jiaotong University. The main domain KGs abroad include the film and television domain graph IMDB [13], the music domain graph MusicBrainz [14], and the geographic domain graph GeoNames [15].

With the development of smart transportation, KG research in the transportation field has gradually become more and more popular. Zhou and Chen [16] combine the urban KG with the deep spatiotemporal convolution neural network to solve the problem of traffic congestion. Zeng et al. [17] use the KG to extract the mine relationships between objects, which models the causal relationships of the equipment failures of the railway trains to ensure the operational safety of the high-speed railway. Muppalla et al. [18] use the KG as an abstraction layer to annotate the traffic incidents collected through various methods. Liu et al. [19] learn the urban traffic characteristics extracted from the urban multisource heterogeneous data and construct a KG to mine the urban mobility patterns. Sun et al. [20] semi-manually construct a microblog traffic event KG by integrating multiple types of open-source data and use such traffic KG and target detection methods to realize the identification of traffic events in microblogs and solve the traffic problems. Liang et al. [21] use the multilevel planning theory to construct an individual travel KG to accurately identify different types of public transport passengers so as to obtain refined public transport travel characteristics and meet the travel needs of different passengers. Zhang et al. [22] integrate the knowledge of interregional flow, events, and weather to enhance the prediction effects of population inflow and outflow in each region of the city.

As an advanced knowledge carrier, the KG has an extremely important position in the portrayal of individual travel. Therefore, we construct an individual travel KG and use it to solve individual travel destination prediction problems.

2.2. Individual Travel Destination Prediction. With the development of urbanization, people's travel patterns are gradually diversified. Understanding human behaviors and modeling individual travel behaviors are helpful to explain some complex socioeconomic phenomena, which is of great value in location-based services, traffic planning, public safety, and so on. Traditional trajectory prediction methods mostly use machine learning methods, such as the hidden Markov model [4], mixed hidden Markov model [23], Bayesian inference [5], and Gaussian mixture model [6]. Based on the research of public transport smart card data, Zhao et al. [24] predict the individual daily travel capacity, and its travel chain is defined as a set of travel start time,

starting point, and destination. Wang et al. [7] extract a variety of features from the subway card swiping data set to predict the travel destination of passengers entering the subway station but not leaving the station. Li et al. [25] improve the prediction effect of the individual travel through clustering the group travel pattern. Wang et al. [26] design a new movement feature, i.e., a time shift tensor, to consider the user's transformation pattern in the time dimension and propose the attention Markov model. Mo et al. [27] analyze the passenger activity pattern based on the public transport card swiping data in Hong Kong and propose an input-output hidden Markov model to predict the time and location of an individual's next trip at the same time.

In recent years, recurrent neural network (RNN) has obtained the excellent performance in modeling sequence data. Wu et al. [28] propose a new robust location prediction model to consider individual preference and social interaction, which alleviates the impact of randomness of location movement and improves the prediction performance. De Brebisson et al. [8] predict the taxi destination by using a multilayer perceptron and a two-way cyclic neural network. Lv et al. [29] regard the trajectory as a two-dimensional image to model the trajectory from different perspectives and apply Convolutional Neural Networks (CNNs) to extract multiscale two-dimensional trajectory features for the accurate destination prediction. Zhang et al. [30] apply Surprisal-Driven Zoneout (SDZ) to RNN, which improves the robustness of the destination prediction model and reduces the training time. Based on the Long Short-Term Memory (LSTM) model, Li et al. [31] combine the extracted depth spatiotemporal features with the original features to predict the taxi destination. Xu et al. [32] use an adaptive attention network to model different extraction features of locations and implement the time gate and the distance gate into LSTM to capture the spatiotemporal relationship between continuous locations.

Although some individual travel destination prediction methods have been proposed, some common problems still seriously affect the prediction effect; for example, it is difficult to grasp the travel intentions of different individuals and handle the sparseness of historical trajectory data of some individuals.

3. Construction of Individual Travel KG

In this section, firstly, we preprocess the subway card swiping data set to clean out the dirty data, the duplicate redundant data, and so on. Then, the individual travel KG is constructed and displayed visually.

3.1. Data Preprocessing. The original data set adopts the passenger card swiping records collected by the Beijing Metro automatic toll collection system in July, August, September, November, and December 2015. Each record contains 11 items, including card ID, subway route, subway station, date, card type, and transaction type. The card ID is the unique identifier of the intelligent transportation card, which is used to identify a unique passenger.

Due to the large volume, the dirty data interference, the missing key items, the coupling of travel records, and other problems in the data set, the card swiping records of each travel individual are scattered in the large data set, which makes it difficult to form a complete travel chain and brings great obstacles in mining the passenger travel patterns. Therefore, the original data must be preprocessed to extract the complete travel chain, so as to build a good data foundation for the construction of individual travel KGs, travel law mining, and travel destination prediction.

First, we remove the repeated card swiping records in the original subway data set and the records with the same card swiping time and the same card swiping station. Moreover, because the traffic card type and other information contribute less to the subsequent destination prediction task, we also filter this information to avoid the interference of redundant information. So, we retain 7 necessary items, including card ID, boarding line, boarding station, alighting line, alighting station, boarding time, and alighting time. In addition, since the research objects are passengers taking the subway as their normal transportation tool, we also remove the passengers whose records are less than the average of 30 card swiping records per month and whose data volume is abnormal. After these operations, 150–400 travel records in five months for each passenger are retained.

3.2. KG Construction. Accurately grasping the individual's travel intention is the major challenge in the task of the individual travel destination prediction, and such travel intention is affected by many factors, for example, time and station. To achieve this purpose, we intend to construct the individual travel KG to accurately analyze the travel individuals and grasp the corresponding travel intention. Therefore, we construct the individual travel KG using the passengers' travel location, time, date, and so on.

In this knowledge, there are five types of entities: card ID, date, travel date attribute (whether working day or not), route, and subway station and times. The corresponding relationships are divided into five categories as shown in Table 1.

We analyzed the travel data of 8000 individuals in five months. Specifically, after data preprocessing, we chose all the historical travel records of 8000 individuals from tens of millions of people. Then, we extracted the travel record knowledge, and the steps are shown in Algorithm 1.

3.3. Visualization of KG. Through the above knowledge extraction steps, we get the structured data and then visually display the graph by neo4j¹. Taking the travel record of Card No. 787931 in July as an example, our constructed graph is shown in Figure 1.

4. Methodology

In this section, we propose a knowledge graph-based enhanced Transformer for the MITD-Pre, namely, KG-Trans. The model framework is shown in Figure 2. Firstly, we identify and extract the travel individuals with similar routes

TABLE 1: Entities and relationships.

Entities	Relationships
Card ID-date	Travel date
Card ID-frequency	Frequency of travel
Date-attributes of travel date	Belong to
Attributes of travel date-route	Travel route
Route-subway station	Boarding/alighting station

Step 1: get the frequency of travel. The frequency of travel of an individual (ID) is the total number of travel records of the ID within a specified time frame in the data set.

Step 2: get the travel date. In the individual travel card swiping record, the travel time is expressed as “year/month/day: hour: minute.”

We only take the date information and ignore the hour information.

Step 3: get the date attribute. For the travel date obtained in Step 2, we check the calendar to determine whether it is a working day. The working day is marked as 1, and the nonworking day is marked as 0.

Step 4: get the origin-destination (OD) records. The storage format of one complete travel record is “ID, boarding time, boarding route, boarding station, boarding time, alighting route, and alighting station.” Then, we extract the travel OD as “boarding line * boarding station - alighting line * alighting station.”

Step 5: get to the subway station. As shown in Step 4, the subway station is expressed in the form of “line * station.”

ALGORITHM 1: Knowledge extraction.

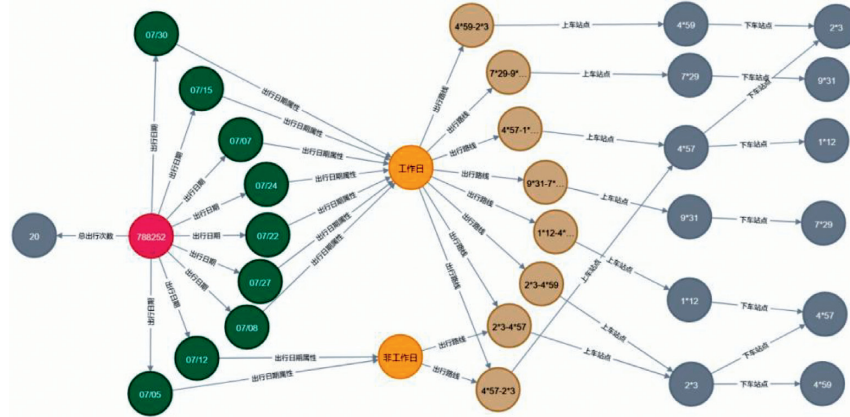


FIGURE 1: Graph display of July travel records of passenger 788252. The front gray node represents the total number of trips per month; the pink nodes represent the individual card ID; the green nodes represent the travel dates; the ginger yellow nodes represent the travel routes; and the behind gray nodes represent the boarding stations and the get-off stations, respectively.

based on the constructed individual travel KG. Then, the same class of data is sent to the Transformer to train the model. Finally, the prediction results are compared to the real values stored in the KG to construct the loss function.

4.1. Relationship Analysis between Travel Individuals Based on KG. Due to the large behavioral differences between individuals and the sparseness of the travel records of some individuals, the accuracy of individual travel destination prediction is seriously affected. In view of this phenomenon, we improve the effectiveness of the individual travel destination prediction through the “group effect.” As for the group here, we define it as “people with similar routes.” For example, for one group of commuters that live in Xierqi, Beijing, and work in Zhongguancun, Beijing, the subway

routes on weekdays are very similar. This paper discovers the individuals having similar routes through analyzing the KG.

KG has many nodes and edges which contain complex information and rich semantics. We aim to find the correlation between nodes in the graph so that we can infer the correlation between the individuals. For example, there are many shared nodes between some passenger routes. As shown in Figure 3(a), we can judge that the two passenger routes are very similar. Some passenger travel routes have few or no shared nodes. As shown in Figure 3(b), it is considered that the similarity between the two passenger routes is relatively weak. In this paper, after analyzing the five-month travel data of 8000 individuals in detail and considering the impacts of classification accuracy on the prediction effect, we constrain the members in the same category satisfy the following rules:

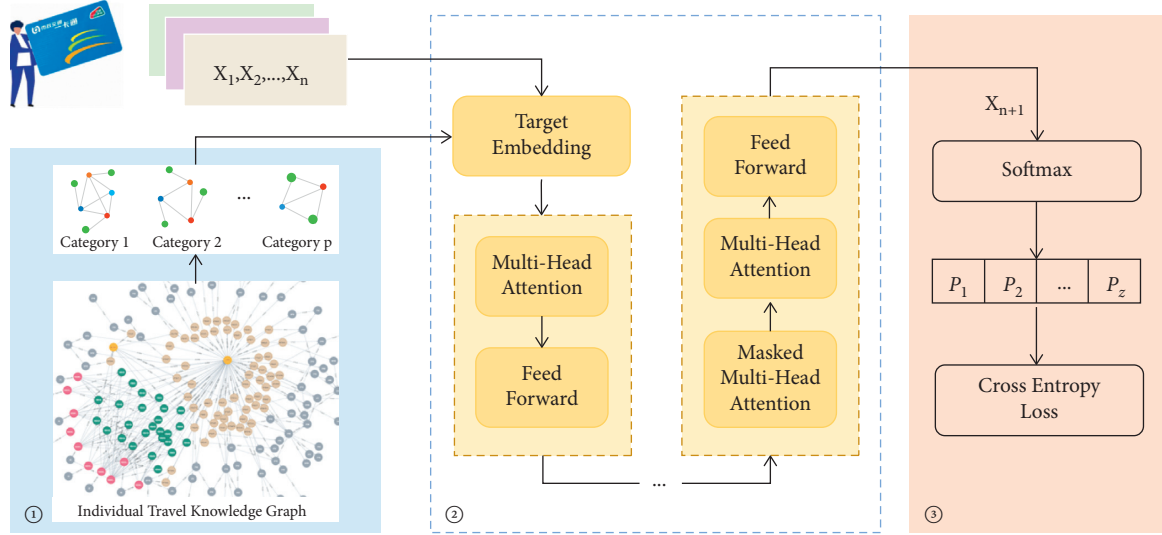


FIGURE 2: KG-Trans model frame diagram. ① The knowledge graph module constructs a multilayer individual travel KG from top to bottom. ② The enhanced Transformer module extracts the dynamic and hierarchical features from the long-term sequential travel trajectory data. ③ The classifier module introduces the cross-entropy loss to constrain the uniqueness of the predicted subway travel station.

- (i) For members in the same category, the difference in travel frequency recorded by card swiping should be less than 50
- (ii) For members in the same category, they should have more than 70% similar routes between the subway stations

Then, on the constructed graph, we first divide the travel records of 8000 people into five length intervals, 150–200, 200–250, 250–300, 300–350, and 350–400 (named group 1 to group 5), according to the travel chain length interval. The node similarity between the members of each length interval is calculated according to the definition of route similarity, and finally, we obtain p classes of travel groups ($p \geq 5$). In these p travel groups, the members of each group are “route similar members” to each other. In this way, by putting each class of members, we classified them into the model for training and we can solve the problem of sparse historical travel routes of some travel individuals according to their similarity.

4.2. Transformer-Based Individual Travel Destination Prediction. The Transformer is an effective method to process the sequence data. Its multihead attention mechanism and stacking layer learn the dynamic and hierarchical characteristics of the sequence data. Therefore, Transformer can predict the traffic flow with the long-term time series and long-term time dependence very well. Considering that the card swiping data of subway passengers have the properties of long-term time series and long-term time dependence, Transformer is naturally selected as an important module in the individual travel prediction model in this paper.

The basic structure of the Transformer used in our model is shown in Figure 2. Its core module is the multihead attention layer. Firstly, the input of the model is the card swiping records of passengers with similar routes, i.e.,

$\mathcal{R} = \{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_m\}$, where m denotes the number of passengers and $\mathcal{X}_i = \{x_0, x_1, \dots, x_n\}$ represents the historical trajectory sequence of the i -th passenger.

The multihead attention (MH) layer adopts different linear mappings to project the input sequence elements to the query, key, and value, i.e., a tuple (Q , K , and V). The output is the corresponding weighted sum of values. The weight assigned to each value is calculated through the compatible functions of the query and the corresponding key. The application of attention can be expressed as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (1)$$

where $\text{Attention}()$ calculates the attention of the input data and $\text{softmax}()$ is an activation function. To establish a single-head attention module, each node can have three subspaces, namely, the queries subspace $Q \in \mathbb{R}^{N \times d_k}$, key subspace $K \in \mathbb{R}^{N \times d_k}$, and value subspace $V \in \mathbb{R}^{N \times d_k}$, where d_k is the dimension of queries, keys, and values.

In the global encoder, the input features are projected into the high-dimensional subspace and the learnable mapping is realized through the feedforward neural network, which can be expressed as

$$\begin{cases} Q = XW_q \\ K = XW_k, \\ V = XW_v \end{cases} \quad (2)$$

where X is the input feature. W_q , W_k , and W_v are the learnable parameters.

The multihead attention network uses h feedforward neural networks to linearly project Q , K , and V , which achieves a multihead mechanism. In this case, the model can pay much attention to the information of different representation subspaces from different stations, which can be expressed as

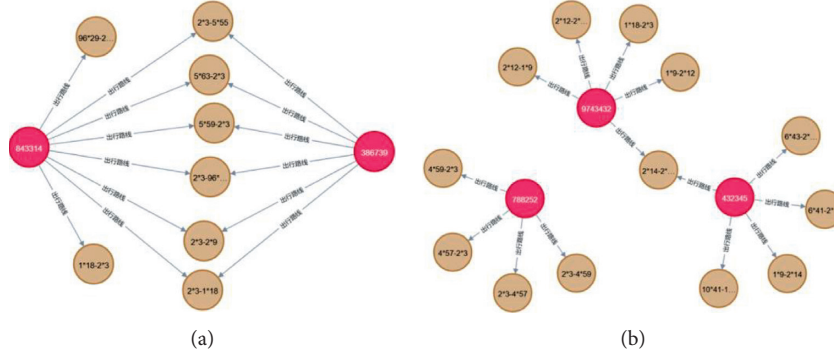


FIGURE 3: Route similarity among different travel individuals: (a) the route similarity is strong; (b) the route similarity is weak. The ginger yellow nodes represent the travel routes, and the pink nodes represent the individual card IDs.

$$\text{MH}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^o, \quad (3)$$

where $\text{MH}()$ calculates the multihead attention of the input data, $\text{Concat}()$ concatenates the input data, and the i -th head head_i can be expressed as

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \quad (4)$$

where the mapping matrices are $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_v}$, and $W_i^o \in \mathbb{R}^{h d_v \times d_{\text{model}}}$, respectively. And, d_{model} is the dimension of keys, values, and queries of single-head attention, $d_k = d_v = d_{\text{model}}/h$.

The final prediction result of our model is

$$x_{n+1} = \text{MH}(Q, K, V), \quad (5)$$

where x_{n+1} is the travel destination of an individual.

Finally, we use the softmax layer to give each station a probability P :

$$P(x_i) = \frac{e^{x_i}}{\sum_z e^{x_z}}, \quad (6)$$

where x_i is the feature representation of i -th station and z is the total number of stations. The station with the largest probability is the final prediction result of the proposed model.

4.3. Loss Function. The individual travel destination prediction problem is different from the previous OD prediction and taxi destination prediction problems. The performance of OD or taxi destination prediction tasks is evaluated by measuring the error of the predicted longitude and latitude. The individual destination prediction results are estimated by only correct (1) or incorrect (0). Therefore, the cross-entropy loss function was selected as the loss function in this paper.

Through the current card swiping records of a travel individual, we want to predict the travel destination of the individual. This problem can be regarded as a probability problem in which we give a probability to each station and the station with the largest probability is the destination prediction result. Cross-entropy mainly reflects the distance

between the actual output (probability) and the expected output (probability). Furthermore, the smaller the value of cross-entropy, the closer the two probability distributions are. Assuming that the probability distribution p is the expected output, the probability distribution q is the actual output, and the cross-entropy $H(p, q)$ measures the distance between p and q , we have

$$H(p, q) = - \sum_{i=1}^n (p(x_i) \log q(x_i) + (1 - p(x_i)) \log(1 - q(x_i))). \quad (7)$$

During training the deep learning network, given the input data and labels, the real probability distribution $p(x)$ is determined. Therefore, the formula of the cross-entropy commonly used in deep learning is formulated as follows:

$$H(p, q) = - \sum_{i=1}^n p(x_i) \log(q(x_i)). \quad (8)$$

5. Experimental Results

5.1. Experiment Settings. In this paper, the proposed destination prediction model is compared with the five traditional prediction methods including Markov, LSTM, GRU, CNN, and FNN.

Markov [4]: the Markov model is a statistical analysis model, which is widely used in speech recognition, automatic part of speech tagging, sequence classification, sequence prediction, and other applications.

LSTM [33]: Long Short-Term Memory (LSTM) can perform better in longer sequences, which is a special version of RNN. The problems of gradient disappearance and gradient explosion in the process of long sequence training are solved.

GRU [34]: the Gated Recurrent Unit (GRU) is an effective variant of the above LSTM network. It has a simpler structure and better effects than the above LSTM network and still solves the long-term dependency problem in RNN. Therefore, it is an important manifold network at present.

CNN [35]: The Convolutional Neural Network (CNN) is a feedforward neural network. For some sequence

processing problems, the effect of the one-dimensional Convolutional Neural Network is comparable to that of the RNN, while the computational cost is usually much lower.

FNN [36]: the feedforward neural network (FNN), also known as a multilayer perceptron (MLP), contains multiple fully connected hidden layers. The complex mapping from input space to output space is realized by aggregating multiple simple nonlinear functions.

We conducted the whole experiments on a GPU workstation, containing a 2080TI GPU with 11G memory. During the training process, the number of epochs is set to 200; the batch size is set to 32; the learning rate is set to 0.001; the number of self-attention heads in the Transformer is set to 8; the dimension of the input vector is set to 6; and the dropout rate is set to 0.3. All experiments use 3 historical observations to predict the next point in this paper.

5.2. Evaluating Indicator. To fairly estimate the experiment performance, we refer to the evaluation indicator Hit@n that is exploited in the KG representation learning TransE [37]. During the testing procedure, the prediction results are arranged from large to small. It determines whether the first n results contain the correct options. If it contains the correct options, we increase the hit value by 1; otherwise, it processes to the next cycle. In other words, we do not require the first one value to be right (Hit@1 except), as long as there is the correct result in the first n results. The final accuracy ACC is calculated as follows:

$$\text{ACC} = \frac{\text{hit}}{\text{hit} + \text{FP}}, \quad (9)$$

where hit is the number of travel stations that are predicted correctly and FP is the number of incorrectly predicted results.

5.3. Prediction Performance of KG-Trans. Table 2 shows the destination prediction results of above five compared methods on the card swiping data set of Beijing Metro. The data set is described in Section 3.1 in detail. It can be seen that our KG-Trans is obviously superior to other methods in all indicators. Group 1–Group 5 in the experimental data set are distributed from less to more according to the number of travel records of one individual (the detailed description is in Section 4.1). The experimental results show that the prediction effect of KG-Trans is more accurate as the number of travel records increases.

For the destination prediction problem solved in this paper, our data set has a long travel sequence of passengers. Although LSTM can mine the long-term sequence characteristics hidden in the data, the LSTM gradient will disappear and affect the prediction results when the sequence length exceeds a certain limit. Moreover, the passenger card swiping data are the sequence data, but the time intervals between records in the sequence do not have the obvious regularity. For the feature learning of such data, LSTM still has difficulty to capture the time correlation.

In Table 2, we can see that the prediction effect of GRU is the worst. We believe the reason is that GRU ignores the

TABLE 2: Experimental results.

		Group 1	Group 2	Group 3	Group 4	Group 5
Hit@3	LSTM	0.383	0.386	0.403	0.490	0.404
	GRU	0.159	0.210	0.128	0.169	0.128
	CNN	0.328	0.578	0.317	0.481	0.469
	FNN	0.547	0.328	0.640	0.688	0.779
	Markov	0.386	0.466	0.489	0.550	0.598
Hit@5	KG-trans	0.572	0.583	0.689	0.731	0.790
	LSTM	0.585	0.538	0.416	0.750	0.689
	GRU	0.309	0.286	0.203	0.258	0.148
	CNN	0.372	0.588	0.426	0.492	0.476
	FNN	0.602	0.464	0.758	0.762	0.794
Hit@10	Markov	0.427	0.522	0.527	0.597	0.640
	KG-trans	0.611	0.760	0.767	0.809	0.810
	LSTM	0.633	0.605	0.470	0.755	0.796
	GRU	0.618	0.433	0.316	0.456	0.160
	CNN	0.450	0.608	0.433	0.875	0.487
Hit@10	FNN	0.672	0.516	0.773	0.883	0.805
	Markov	0.486	0.589	0.600	0.659	0.690
	KG-trans	0.733	0.775	0.792	0.947	0.929

middle layer compared to LSTM. Although this operation reduces parameters and prevents over-fitting risk, the middle layer plays a key role in the long-term sequence feature extraction. Therefore, the GRU effect receives the worst performance in our passenger travel prediction scenario.

The prediction effect of CNN is similar to that of LSTM. The receptive field of CNN has a fixed size, and the processing effect of the local information is very prominent. It can capture some local specific features, while the capture ability of global features is worse than LSTM.

In addition to our model, FNN has the best experimental effects. FNN is a fully connected network, and the computational complexity is very large, which results in low training efficiency. In addition, FNN can only learn the high-order combined features while the low-order features are not modeled in the model.

Since the prediction effect of Markov heavily depends on the data of the previous time, the performance of the characteristic learning of long-term time series is poor, so the prediction effect of individual travel destinations is not good.

Compared with the above methods, with the help of the accurate portrait analysis of passengers in our individual travel KG, we can identify and extract the route similar passengers and put them into different in-depth learning models for training. Such KG solves the problem of the sparse individual travel trajectory by exploiting the characteristics of similar passenger routes. In addition, the Transformer has the multihead self-attention mechanism and the stacking layer, which has stronger structural flexibility and captures a wider range of time correlation. The Transformer achieves a very good prediction effect for the individual travel data series with the long-term time series and the long-term time dependence.

6. Conclusion

In this paper, a knowledge graph-based enhanced Transformer for the metro individual travel destination prediction method is proposed. The method of constructing individual travel KG is used to accurately analyze the travel individuals. And then, the Transformer's outstanding sequential learning ability is used to capture the sequence information in the individual travel chain. The test results on the AFC card swiping data set of Beijing Metro show that our method can well learn the regularity and characteristics of passenger travel records, which is greatly improved compared with previous studies. In future work, our model should also consider the influence of more external factors (the weather, the road network, and traffic events) and build more levels of traffic KG to improve the prediction accuracy. Moreover, in terms of individual travel destination prediction, in addition to the next travel location, the next travel time will also be the focus of our next research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research project was partially supported by the National Natural Science Foundation of China under Grant nos. 62072015, U19B2039, and U1811463 and National Key R&D Program of China under Grant no. 2018YFB1600903.

References

- [1] X. L. Wei, Y. Zhang, Y. Wei et al., "Metro passenger-flow representation via dynamic mode decomposition and its application," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [2] J. Wang, Y. Zhang, Y. Wei, Y. Hu, X. Piao, and B. Yin, "Metro passenger flow prediction via dynamic hypergraph convolution networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 12, pp. 7891–7903, 2021.
- [3] Y. Wang, Y. Zhang, X. Piao, H. Liu, and K. Zhang, "Traffic data reconstruction via adaptive spatial-temporal correlations," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1531–1543, 2018.
- [4] D. Lian, X. Xie, V. W. Zheng, N. J. Yuan, F. Zhang, and E. Chen, "Cepr: a collaborative exploration and periodically returning model for location prediction," *ACM Transactions on Intelligent Systems and Technology*, vol. 6, no. 1, pp. 1–27, 2015.
- [5] J. Krumm and E. Horvitz, "Predestination: inferring destinations from partial trajectories," in *Ubiquitous Computing, 8th International Conference*, Berlin, Germany, 2006.
- [6] J. Wiest, M. Hoffken, U. Kreßel, and K. Dietmayer, "Probabilistic trajectory prediction with Gaussian mixture models," in *Intelligent Vehicles Symposium (IV)*, pp. 141–146, Madrid, Spain, 3–7 June 2012.
- [7] H. Wang, J. Zhao, K. Ye et al., "A destination prediction model for individual passengers in urban rail transit," *International Conference on High Performance Big Data and Intelligent Systems*, pp. 1–6, 2020.
- [8] A. De Brébisson, É. Simon, A. Auvolat, P. Vincent, and Y. Bengio, "Artificial neural networks applied to taxi destination prediction," *Computer Science*, 2015.
- [9] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge," in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, pp. 1247–1250, New York, NY, USA, 2018.
- [10] A. Sören, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A nucleus for a web of open data," *Semantic Web*, vol. 4825, pp. 722–735, 2007.
- [11] Y. Wang, Y. Jia, D. Liu, X. Jin, and X. Cheng, "Open web knowledge aided information search and data mining," *Journal of Computer Research and Development*, vol. 52, no. 02, pp. 456–474, 2015.
- [12] Y. Ma and G. Qi, "An analysis of data quality in DBpedia and Zhishi.me," *Linked Data and Knowledge Graph*, vol. 406, pp. 106–117, 2013.
- [13] W. Peng, S. Li, G. Sun et al., "RC-NVM: Enabling symmetric row and column memory accesses for in-memory databases," in *IEEE International Symposium on High Performance Computer Architecture*, pp. 518–530, Vienna, Austria, 24–28 Feb. 2018.
- [14] A. Swartz, "MusicBrainz: A semantic web service," *IEEE Intelligent Systems*, vol. 17, no. 1, pp. 76–77, 2002.
- [15] B. Regalia, K. Janowicz, G. Mai, D. Varanka, and E. L. Uesry, "GNIS-LD: Serving and visualizing the geographic names information system gazetteer as linked data," in *European Semantic Web Conference*, Cham, Switzerland, 2018.
- [16] G. L. Zhou and F. Chen, "Urban congestion areas prediction by combining knowledge graph and deep spatio-temporal convolutional neural network," in *2019 4th International Conference on Electromechanical Control Technology and Transportation*, pp. 105–108, Guilin, China, 26–28 April 2019.
- [17] Y. Zeng, Y. Qin, D. Liu, Y. Fu, M. Gong, and X. Zhang, "Railway train device fault causality model based on knowledge graph," in *2020 International Conference on Sensing, Diagnostics, Prognostics, and Control*, pp. 385–390, Beijing, China, 5–7 Aug. 2020.
- [18] R. T. Muppalla, S. Lalithsena, T. Banerjee, and A. Sheth, "A knowledge graph framework for detecting traffic events using stationary Cameras.WebSci '17," in *ACM Web Science Conference*, pp. 431–436, New York, NY, USA, 2017.
- [19] J. Liu, T. Li, S. Ji et al., "Urban flow pattern mining based on multi-source heterogeneous data fusion and knowledge graph embedding," *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [20] X. R. Sun, Y. Meng, and W. L. Wang, "Identifying traffic events from weibo with knowledge graph and target detection," *Data Analysis and Knowledge Discovery*, vol. 48, no. 12, pp. 140–151, 2020.
- [21] Q. Liang, J. C. Weng, and P. F. Lin, "Public transport commuter identification based on individual travel graph," *Journal of Transportation Systems Engineering and Information Technology*, vol. 18, no. 002, pp. 100–107, 2018.
- [22] J. Zhang, Z. Yu, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," *AAAI*, vol. 31, 2016.
- [23] J. Ye, Z. Zhu, and H. Cheng, "What is your next move: User activity prediction in location-based social networks," in

Proceedings of the SIAM International Conference on Data Mining, Austin, TX, USA, 2013.

- [24] Z. Zhao, H. N. Koutsopoulos, and J. Zhao, "Individual mobility prediction using transit smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 89, no. APR, pp. 19–34, 2018.
- [25] F. Li, Q. Li, Z. Li, Z. Huang, X. Chang, and J. Xia, "A personal location prediction method based on individual trajectory and group trajectory," *IEEE Access*, vol. 7, pp. 92850–92860, 2019.
- [26] H. Wang, Y. Li, D. Jin, and Z. Han, "Attentional Markov model for human mobility prediction," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 2213–2225, 2021.
- [27] B. Mo, Z. Zhao, H. N. Koutsopoulos, and J. Zhao, "Individual mobility prediction in mass transit systems using smart card data: an interpretable Activity-based hidden Markov approach," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [28] R. Wu, G. Luo, Q. Yang, and J. Shao, "Learning individual moving preference and social interaction for location prediction," *IEEE Access*, vol. 6, pp. 10675–10687, 2018.
- [29] J. M. Lv, Q. Li, Q. Sun, and X. Wang, "T-CONV: A convolutional neural network for multi-scale taxi trajectory prediction," *Computer Vision and Pattern Recognition*, pp. 82–89, 2016.
- [30] G. Zhang, Y. Li, L. Zhang, Q. Fan, and X. Li, "Taxi travel destination prediction based on SDZ-RNN," *Computer Engineering and Applications*, vol. 54, no. 6, pp. 143–149, 2018.
- [31] Y. Li, S. Cui, L. Zhang, B. Liu, and D. Song, "Taxi destination prediction with deep spatial-temporal features," in *International Conference on Communications, Information System and Computer Engineering*, pp. 562–565, Beijing, China, 14–16 May 2021.
- [32] J. Xu, J. Zhao, R. Zhou, C. Liu, P. Zhao, and L. Zhao, "Predicting destinations by a deep learning based approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 2, pp. 651–666, 2021.
- [33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [34] K. Cho, B. V. Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *Computer Science*, 2014.
- [35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.
- [36] G. Bebis and M. Georgiopoulos, "Feed-forward neural networks," *IEEE Potentials*, vol. 13, no. 4, pp. 27–31, 1994.
- [37] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, *Translating Embeddings for Modeling Multi-Relational Data*, Curran Associates Inc, NY, United States, 2013.