Short-term Traffic Flow Prediction Based on the SGA-KGCN-LSTM Model

Qingrong Wang, Xiaohong Chen, Changfeng Zhu, Wei Chai

***Abstract*—Traffic volume forecast is the key to alleviating traffic congestion. However, the relationship between traffic data and outside factors makes the problem more complex. Existing traffic flow forecasting studies seldom consider the relationship between traffic data and outside factors. Therefore, we propose SGA-KGCN-LSTM to solve this problem. The model combines Savitzky-Golay (SG) filter, Knowledge Graph (KG), Graph Convolution Network (GCN), Long Short-Term Memory (LSTM), and Self-Attention Mechanism. Firstly, the SG filter can be employed to reduce the noise of traffic volume data. Then, aiming at the relationship between traffic data and outside factors, the KG theory is introduced, and the knowledge representation is applied to get the embedding of relevant knowledge. Secondly, new road features are obtained by integrating embedded information and traffic flow characteristics. GCN can be employed to acquire the spatial characteristic of traffic stream, and LSTM can be employed to extract the temporal characteristics of traffic stream. Finally, the input feature information is given enough weight by self-attention mechanism. The outcome is subsequently acquired by utilizing the fully connected layer to attain the ultimate consequence. The trajectory data of the Luohu taxi in Shenzhen are used in the experiment. The experimental consequence indicate that the SGA-KGCN-LSTM has high forecasting precision compared with the benchmark model and the ablation experiment.**

***Index Terms*—Knowledge Graph, Long short-term memory, Savitzky-Golay filter, Self-attention mechanism**

I. INTRODUCTION

IN recent years, the rising demand for travel has made it increasingly challenging to meet people's transportation needs. The matter of traffic congestion has reached alarming levels in recent times, and addressing this issue has become a top priority for modern society. The Intelligent Transportation System (ITS) has the underlying to offer effective route recommendations to travelers and help alleviate traffic congestion [1-2]. Short-term traffic volume forecast is a crucial constituent part of ITS, and study of short-term traffic volume forecast theory holds immense theoretical and practical significance in promoting the implementation of ITS and mitigating traffic congestion [3].

During the early phase, statistical models were used to process historical traffic data using mathematical statics, which mainly included the History Average model (HA), Kalman filtering model, ARIMA, etc. Although the HA model has a simple structure and fast calculation speed, it cannot handle complicated traffic situations and is only appropriate for static forecasting systems with low precision demands [4]. In literature [5], a hybrid prediction method combining DFT and SVR was proposed. DFT was used to identify the common trend in traffic flow data by setting an appropriate threshold. SVR was then applied to predict the residual sequence by extrapolating the historical trend limit. Literature [6] introduced a traffic flow prediction theory that combines parameterization and bootstrap-up technology. The combination of bootstrap-up method with the traditional ARIMA model improves the precision of prediction while keeping theoretical consistency.

Despite their usefulness, the earlier theoretical methods mentioned above have certain limitations when it comes to solving irregular and sudden traffic conditions. For example, the ARIMA model is more suited to handle linear data, and as a result, it has some limitations when dealing with nonlinear data, which can affect the accuracy of traffic flow forecasting. Therefore, there is a need for more advanced techniques that can effectively handle the complexities and nonlinearity of traffic data to boost the exactness of short-term traffic volume forecast.

To overcome the drawbacks of the earlier prediction methods, researchers have turned to machine learning techniques for forecasting traffic flow. In literature [7], a precise traffic flow forecasting model that employs Support Vector Machines (SVM) for multi-step forecasting was introduced. Notwithstanding the benefits of machine learning in handling intricate traffic data, capturing long-term memory features of traffic flow sequences remains challenging due to its limited generalization ability, which can ultimately impact the accuracy of machine learning prediction [8].

The evolution of artificial intelligence in recent years resulted in the progress of deep learning, which can effectively address some of the limitations of traditional machine learning techniques. In literature [9], existing research methods overlooked the time and space attributes of traffic information, and used Convolutional Neural Network (CNN) model for short-term traffic volume forecast. In addition, literature [10] presented an approach relies on LSTM for handing irregular sampling and data loss of traffic stream data. In [11], a time-information-enhanced LSTM model was put forward to forecast traffic stream for a single section. Additionally, deep learning was employed in traffic flow research by [12] for conducting time series analysis. Another approach was proposed in [13], which used an

Manuscript received January 17, 2023; revised July 03, 2023.

This work was supported in part by the National Natural Science Foundation of China (No. 71961016, 72161024), "Double-First Class" Major Research Programs, Educational Department of Gansu Province (No. GSSYLXM-04).

Q. R. Wang is a professor of Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 329046272@qq.com).

X. H. Chen is a postgraduate student of Lanzhou Jiaotong University, Lanzhou 730070, China. (Corresponding author to provide phone: 15095408625; e-mail: 2937325132@qq.com).

Ch. F. Zhu is a professor of Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: cfzhu003@163.com).

W. Chai is a postgraduate student of Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 403808465@qq.com).

auto-encoder LSTM network for traffic volume prediction. In [14], an improved genetic algorithm was proposed to ameliorate the LSTM prediction model, taking into account the strong time-series correlation features of short-term road traffic stream.

The raw data can be contaminated by noise, which can have an impact on the model's predictions. To address this issue, [15] introduced a data denoising approach that combines empirical mode decomposition and wavelet to deal with outliers in the original data, and used the LSTM model to realize traffic flow forecasting. Similarly, [16] adopted wavelet filtering, moving average filtering, and Butterworth filtering to denoise traffic stream data and obtained noiseless traffic stream data. The experiments conducted in these studies demonstrate that denoising the original data can effectively enhance the precision of traffic volume forecasting. However, some scholars have also explored the strategy of adding noise to the data to boost the model's generalization capacity. It is worth noting that excessive noise can lead to data jitter and other phenomena.

The Savitzky-Golay (SG) filter is a widely used signal processing technique that can eliminate high-frequency noise while maintaining the low-frequency elements of a signal. In [17], an improved network traffic prediction method called ST-LSTM was designed by combing the SG filter, Temporal Convolutional Networks (TCNs), and LSTM. Firstly, the SG filter was applied to remove noise from the data. Then, TCN was used to analyze the time characteristics of the data. Finally, LSTM was applied to model the long-term dependence in the time series. Similarly, [18] used an SG filter to smooth water quality time series and proposed a comprehensive prediction method that combines SG filtering with an encoder-decoder neural network based on LSTM.

Traffic flow forecasting is influenced by complex spatial-temporal factors, and a single model may not be able to capture all of these factors simultaneously. Therefore, researchers have proposed combined prediction models. For example, [19] presented a KNN-LSTM model for traffic stream forecast that incorporates both spatiotemporal characteristics of traffic stream. [20] proposed a short-term traffic stream forecasting model that takes into account traffic flow's correlation between time and space. [21] introduced a model based on spatiotemporal multi-graph convolution networks that simultaneously captures the spatial, temporal, and semantic relationships between different global features of the road network for predicting traffic flow. Similarly, [22] utilized CNN and LSTM to forecast traffic stream by capturing the spatiotemporal characteristics of traffic stream data. [23] developed a multi-feature fusion model based on deep learning. [24] presented a time-graph convolution network model that uses GCN to acquire the traffic stream's spatial characteristics and GRU to capture the temporal relationship.

The attention mechanism can enhance the precision of deep learning models by adjusting the weight according to the importance of different inputs. With the increase in the longness of the time series, the memory ability of LSTM may decline. To address this issue, researchers have introduced the attention mechanism to accurately locate the relationship between the local and the whole before each prediction. However, previous research on traffic flow forecasting rarely consider its application in this field. Therefore, [25] introduced the attention mechanism theory and used LSTM-RNN model to forecast short-term traffic stream considering the advantages of LSTM and RNN models. Similarly, In [26], a model for forecasting traffic volume found on a spatiotemporal convolution graph attention network was proposed. [27] proposed a mixed deep learning short-term traffic volume forecasting model based on the Conv-LSTM network, incorporating the attention mechanism. [28] used spatiotemporal attention to acquire the spatial characteristics between road sections and the time dependence between time steps. [29] proposed a new spatiotemporal GCN model found on attention to address the problem of traffic volume forecast. Finally, [30] proposed a multi-step forecasting model that combines attention CNN and short and long-time memory networks.

Although the attention mechanism can boost the performance of models and mitigate the issue of memory loss, it heavily relies on external information. The self-attention mechanism is more effective in capturing relevant information within features and reducing their reliance on external data. In reference [31], a SA-LSTM model was proposed grounded on the self-attention mechanism. [32] introduced a complementary model for missing traffic volume data. The model integrates the self-attention mechanism, auto-encoder, and generative adversarial network to form the self-attention generative adversarial complement network.

Exterior factors, such as extreme weather and public holidays, can also influence the traffic stream. However, a large proportion of the current research have overlooked the effect of these exterior factors on traffic stream. In reference [33], a short-term traffic volume forecasting theory that integrates multiple factors was proposed. [34] took into account the impact of adverse weather conditions on traffic stream and utilized an artificial intelligence-based short-term traffic prediction model for meteorological conditions in smart cities. [35] utilized graph wavelet transform and adaptive matrix to extract local and global traffic flow features and proposed a combined prediction model that incorporates external attributes. To account for weather conditions, reference [36] created a model for short-term traffic volume forecast that relies on the 1DCNNLSTM-Attention networks. In reference [37], a spatiotemporal graph convolutional network (GCN) that incorporates attention and external factors was proposed for multi-step traffic flow forecasting.

In conclusion, although the above methods have carried out certain research on short-term traffic volume forecast. However, there are still some questions that need to be addressed, such as:

(1) The traditional models used for short-term traffic volume forecast are relatively uncomplicated. While they can forecast the changing tendency of future traffic volume to some extent, they are not capable of handling complex traffic conditions. Machine learning approaches can compensate for the limitations of traditional models, but their generalization ability is not high. The limitations of the aforementioned models can be addressed by deep learning models. However, with the growing complexity of the model, it becomes more expressive and requires a larger amount of information to

store, which can lead to the matter of information overload.

(2) Most of the existing research on short-term traffic volume forecast has not considered the relatioship between traffic data and outside factors. KG can clearly and precisely express the semantic connections among attributes and transactions. Within the domain of traffic volume forecasting, the use of KG can represent the aforementioned correlation. Without KG, external factors can only be incorporated into traffic volume prediction, and the relationship between them cannot be properly obtained.

(3) The data collected is susceptible to noise. If noise reduction is not performed on the data, it can significantly impact the speed of convergence of the data and even the precision of the model. Owing to the significant variations of time series data and the overlap between useful signals and noise, it can be challenging to separate the signal from the noise using traditional methods. Savitzky-Golay (SG) filter is a technique that can more effectively preserve the information related to signal changes while smoothing out noise. It is particularly suitable for situations where data changes need to be closely monitored.

To address this issue, firstly, KG theory is introduced, and the embedding of associated knowledge is obtained by using knowledge representation. On the other hand, the initial data is denoised by the SG filter. Secondly, knowledge fusion cells are employed to fuse the embedded information with the denoised traffic flow features, creating new road section features. The GCN-LSTM approach is utilized to acquire the spatiotemporal characteristics of traffic stream data. Finally, the self-attention mechanism is applied to determine the relationship between input features. Then the idea of a probability distribution is adopted to assign sufficient weight to the input characteristics, enhancing the precision of traffic volume forecasting.

II. Correlation Analysis And Feature Information Fusion of External Traffic Attributes Based on KG

A. Traffic Flow Attribute Analysis

1) Analysis of the spatiotemporal features of traffic stream

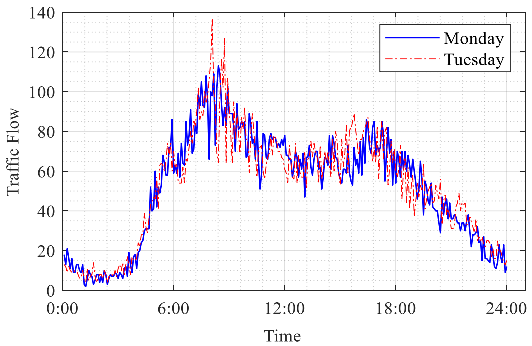


Fig. 1. The traffic flow distribution of the same section between adjacent days

Traffic flow exhibits spatiotemporal features. The accumulation and dispersal of vehicles on a road segment occur gradually over time. The traffic stream in the upcoming moment is influenced by the preceding moment in time. Spatially, the traffic flow in adjacent sections tends to be similar [38]. Additionally, due to factors such as work schedules, the traffic flow from one workday to the next workday is often similar. The traffic flow between two adjacent days also exhibits similar trends.

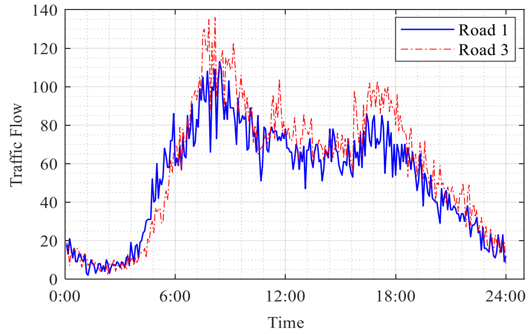


Fig. 2. The traffic flow distribution on different sections on the same day

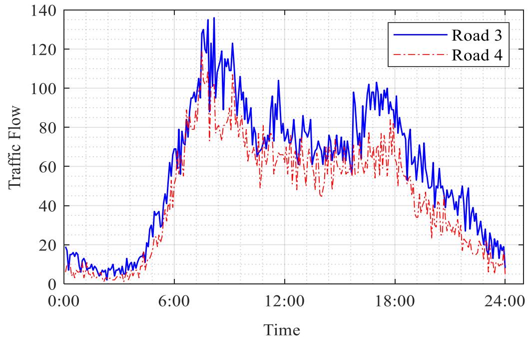


Fig. 3. The traffic flow distribution on adjacent sections on the same day

The data presented in the figure above sourced from the Performance Measurement System (PeMS), which is sponsored by the California Department of Transportation. As shown in Fig. 1, the traffic flow on workdays exhibits obvious peak periods in the morning and evening, which is closely related to people's commuting habits of working from nine to five. Moreover, on the same sections, the traffic stream distribution between adjacent days exhibits similarities. Fig.2 manifest the traffic stream distribution on different road segments on the same day, and we can see that the traffic stream on different sections on the same day exhibits similar trends. Fig. 3 shows the traffic stream distribution on neighboring road sections on the same day, and it can be seen that they also exhibit similarities.

2) Analysis of external attributes of traffic flow

a. Weather factors.

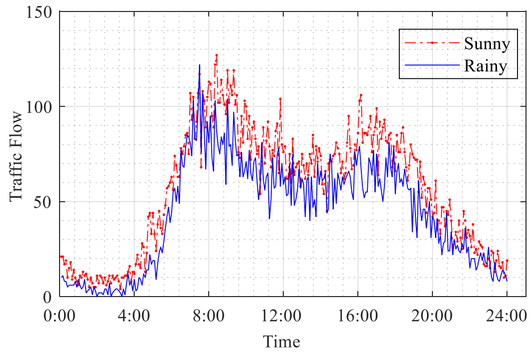


Fig. 4. The traffic flow distribution under different weather conditions

The weather conditions have an impact on the traffic stream. If the weather is sunny and the traffic stream on the road is relatively large, if the weather is drizzle, moderate rainfall, or intense rainfall, then the traffic stream will reduce sharply. Fig. 4 compares the distribution of traffic flow under two different weather conditions, sunny and light rain. We can see that the traffic stream on the section is higher when the weather is sunny.

b. POI factors.

In addition to weather, traffic flow is also influenced by POI factors. We analyzed the impact of Accommodation Services (AS), Medical Services (MS), Life Services (LS), Shopping Services (SS), and restaurant services (CS) on traffic flow. It is observed that the shopping and catering services on Road 2 are more numerous than on the other two sections. Therefore, the traffic stream on this section is higher than that on other two sections.

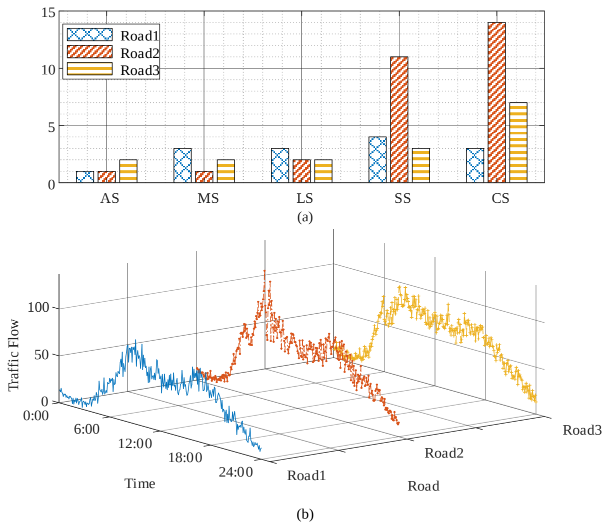


Fig. 5. The impact of POI factors on traffic flow

B. External Attributes and Their Correlation Analysis Based on KG

1) Knowledge Graph

Knowledge Graph (KG) is a semantic network designed to reveal the relation between all things [39]. It can integrate data from different sources. KG uses Resource Description Framework (RDF) to store data in a triple (head, relationships, tail) way. The knowledge graph construction process is shown in Fig. 6.

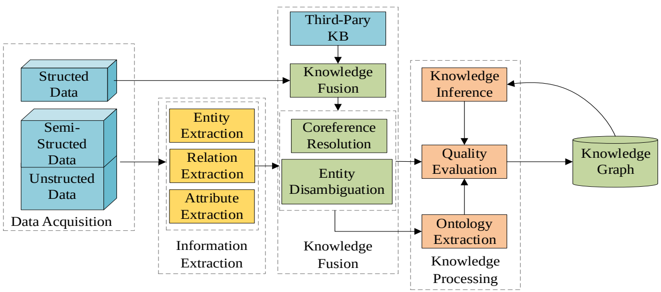


Fig. 6. Knowledge graph construction process

2) Knowledge Representation

The traditional knowledge representation mostly uses representation based on entity relationships. By symbolizing the entity and relationship and mapping them to the low-dimensional space, the low-dimensional vector represents the original data sample, which performs well in extracting entity relationships. Traditional knowledge representation is shown in Fig. 7.

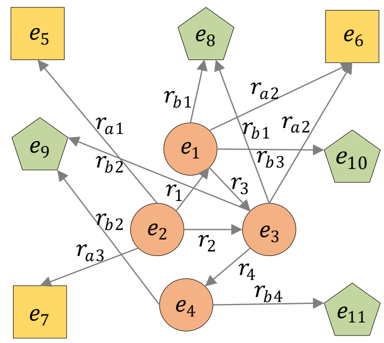


Fig. 7. Traditional knowledge representation method

In the knowledge representation based on entity relation, relation mainly refers to the following two points: First, it represents the entity's attributes. Second, it represents the relationship between entities and entities. However, attribute and relation are two different concepts. Attribute refers to the entity's attribute, while relation refers to the non-semantic relation between entity and entity. Traditional knowledge representation does not distinguish these two concepts.

The relation between the external attributes and the traffic network is many-to-many in the traffic network. When the traditional knowledge represents the modeling of the many-to-many relationship, the accuracy is relatively poor, which is unsuitable for this situation. Therefore, this paper uses KR-EAR [40] to acquire the structure of knowledge between the section and external factors and the semantic information between them. The knowledge representation based on entity-attribute-relationship is shown in Fig. 8. (e_3, r_4, e_4) is relational triple, and (e_4, r_{b4}, e_{11}) represents an attribute triplet.

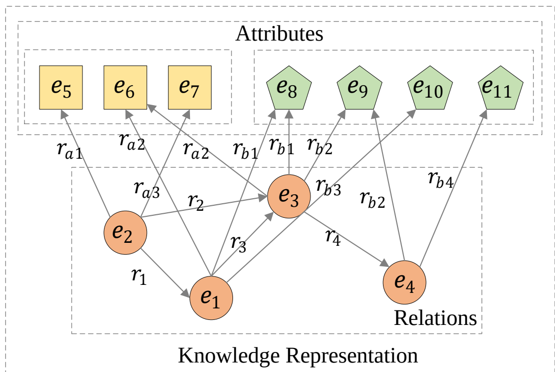


Fig. 8. Representation of KR-EAR

This paper utilizes K= {R, S, A} to denote roads, attributes, and their interrelationships. Specifically, R represents the relational triplet that indicates whether an adjacency relationship exists between two sections:

$$R=\left\{\left(v_i, adj, v_j\right), i, j\in\left\{1,2,\cdots,n\right\}\right\} \tag{1}$$

Where: v_i and v_j represent section i and section j ; adj is the adjacency between v_i and v_j ; n represents the quantity of roads.

The attribute triplet S denotes the corresponding relationship between a road and its attribute:

$$S=\left\{\left(v_i, s_l, s_l^{v_i}\right), i\in\left\{1,2,\cdots,n\right\}, l\in\left\{1,2,\cdots,L\right\}\right\} \tag{2}$$

Where: s_l represents the l class attribute, $s_l^{v_i}$ represents the

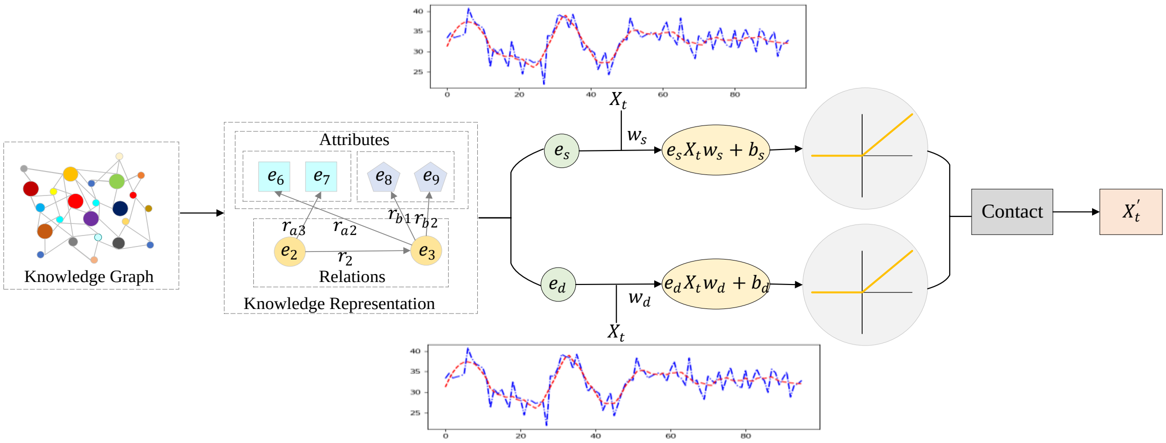


Fig. 9. Structure diagram of KF-Cell

attribute value corresponding to the section v_i, and L represents the overall count of attribute categories.

The co-occurrence relation A between two different attributes is

$$A=\{(s_{l1},s_{l2},p),l_1,l_2\in\{1,2,\cdots,L\}\} \tag{3}$$

Where: p represents the co-occurrence probability of class l_1 attribute and class l_2 attribute.

Use X_E to represent the embedding of entities, relations, and attributes, given their premise, assuming that R and S are independent. The joint probability of R and S is maximized, and the objective function is defined through this operation, namely:

$$P(R,S|X_E)=P(R|X_E)P(S|X_E)= \prod_{(v_i,adj,v_j)\in R} P((v_i,adj,v_j)|X_E) \prod_{(v_i,s_l,s_l^{v_i})\in S} P((v_i,s_l,s_l^{v_i})|X_E) \tag{4}$$

Where: $P((v_i,adj,v_j)|X_E)$ is the contingent probability of relational triples (v_i,adj,v_j). $P((v_i,s_l,s_l^{v_i})|X_E)$ is the conditional probability of attribute triples $(v_i,s_l,s_l^{v_i})$.

$V=\{v_1,v_2,\cdots,v_n\}$ represents a group of road segment entities. The conditional probability of relational triples is produced by TransR [41].

$$P((v_i,adj,v_j)|X_E)=\frac{\exp(g(v_i,adj,v_j))}{\sum_{v_i\in V}\exp(g(\hat{v}_i,adj,v_j))} \tag{5}$$

$$g(v_i,adj,v_j)=-\|v_iM_r+adj-v_jM_r\|_{L_1/L_2}+b_1 \tag{6}$$

Where: $g()$ is the energy function and represents the correlation between the relation and entity pair. L_1 and L_2 represent the L_1 and L_2 norms. M_r is the transfer matrix. b_1 is the offset term.

Relationships between entities and attributes should be extracted using a disaggregated model. In the generic triplet $(v_i,s_l,s_l^{v_i})$, The conditional probability of attribute triples is described as:

$$P((v_i,s_l,s_l^{v_i})|X_E)=\frac{\exp(h(v_i,s_l,s_l^{v_i}))}{\sum_{s_l^{v_i}\in s^v}\exp(h(v_i,s_l,s_l^{v_i}))} \tag{7}$$

$$h(v_i,s_l,s_l^{v_i})=-\|f(v_iw_s+b_s)-E_{s^v}\|L_1/L_2+b_2 \tag{8}$$

Where: f represents a nonlinear function. s^v represents an attribute's value set. E_{s^v} represents attribute value s^v embedded vector, w_s represents linear transformation, b_s and b_2 represents bias.

C. KG-based Integration of Traffic Flow Characteristics and External Attribute Associations

This paper introduces the concept of KG theory and utilizes knowledge representation to embed relevant knowledge. KF-Cell [42] integrates KG-based traffic flow characteristics with external attribute correlation to generate new features for roads. Fig. 9 illustrates the structure of KF-Cell.

Because of the multiformity of external factors, knowledge embedding is segmented into static and dynamic external factors. The formula for calculating KF-Cell is as follows:

$$X_s=Relu(e_sX_tw_s+b_s) \tag{9}$$

$$X_d=Relu(e_dX_tw_d+b_d) \tag{10}$$

$$X_t'=[X_s,X_d] \tag{11}$$

Where: e_s and e_d represent static external factors and dynamic external factors, respectively; X_t is the sectional characteristics observed at the time t; w_s and w_d represent linear transformations; b_d represent offset items; $[\cdot]$represents splicing; X_t' represents the updated section features after the integration of external knowledge at the time t.

III. SHORT-TERM TRAFFIC FLOW PREDICTION MODEL

A. Noise Reduction Strategy Based on SG Filter

The SG filter, also known as a smoothing, is a denoising method that employs local polynomial least square fitting in the time domain. The least-square smoothing is shown in Fig.10.

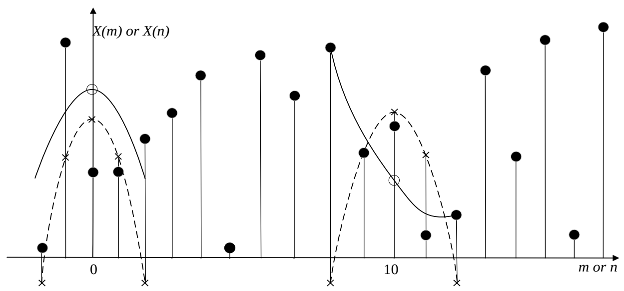


Fig. 10. Least squares smoothing legend

SG filtering methods typically utilize the least-squares means to fit a polynomial to the data within a fixed window length, and subsequently employ the polynomial to estimate the value of the data point in the center of the window. This operation helps to smooth out the noise-contaminated data so that the original data is as free from noise pollution as possible. SG filtering can select different window lengths at any position on the same curve to achieve various smoothing and filtering effects. This technique is particularly advantageous when dealing with time series data, as it can effectively handle sequences with diverse patterns.

SG filtering is commonly used in the smoothing and de-noising of data streams. Its dominating idea is to fit the data in a certain size window by a K-order polynomial. It will remove high band noise while retaining low band noise. The following is the SG smoothing formula:

$$x_{k,smooth} = \frac{1}{H}\sum_{i=-w}^{+w} x_{k+i} h_i \qquad (12)$$

Where: $x_{k,smooth}$ represents the output data after K-order smoothing. $\frac{h_i}{H}$ represents the smoothness coefficient obtained by fitting the polynomial with the least square method, and w represents the number of unilateral points that need to be fitted, as well as x_{k+i} represents the data to be fitted.

Different noises will pollute the collected data. Excessive noise will cause data jitter and other conditions, influencing traffic flow forecasting and thus affecting the model's performance. Due to the large time series fluctuation, useful signal and noise overlap seriously, and it is difficult to achieve effective separation of signal and noise by traditional denoising method. SG filtering method can effectively retain the change information of signal while smoothing, which is more suitable for some occasions that pay attention to the data change. Therefore, this paper proposes a noise-lessened strategy for the original data based on the SG filter.

B. KGCN Model

The conventional CNN model is designed to process data in matrix form, which is suitable for the Euclidean structure. However, the GCN model is tailored to handle data in a topological structure that is more suitable for non-Euclidean structures. Given that traffic networks are composed of irregular non-Euclidean structures, the GCN model is a more appropriate choice for processing such data than the traditional CNN model. The present study utilizes the GCN model to acquire spatial characteristics of traffic stream. Specifically, the representative vector of each section is calculated as follows:

$$gcn\left(H^{(l)},A\right)=\sigma\left[\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right] \qquad (13)$$

Where: σ represents l layer nonlinear activation function. $\tilde{A}=A+I$, A is the adjacent matrix, and I represents the unit matrix. $\tilde{D}$ represents the degree matrix of $\tilde{A}$. $H^{(l)}$ is the output of the L-layer node characteristic nonlinear combination and $W^{(l)}$ represents the l layer's weight matrix.

KG theory is introduced, aiming at the dependency problem between traffic information and external attributes, and knowledge representation is used to obtain the embedding of correlation knowledge. Then KF-Cell is applied to fuse the embedded information with the traffic flow features after noise reduction. Take the result of the fusion and adjacency matrix as inputs to the GCN model. The structure diagram of KGCN is shown in Fig. 11.

C. LSTM Model

Recurrent Neural Networks (RNN) suffer from the matter of "long-term dependence," where the gradient vanishes or explodes when processing long time series data. To overcome this issue, the LSTM model has been favored by many

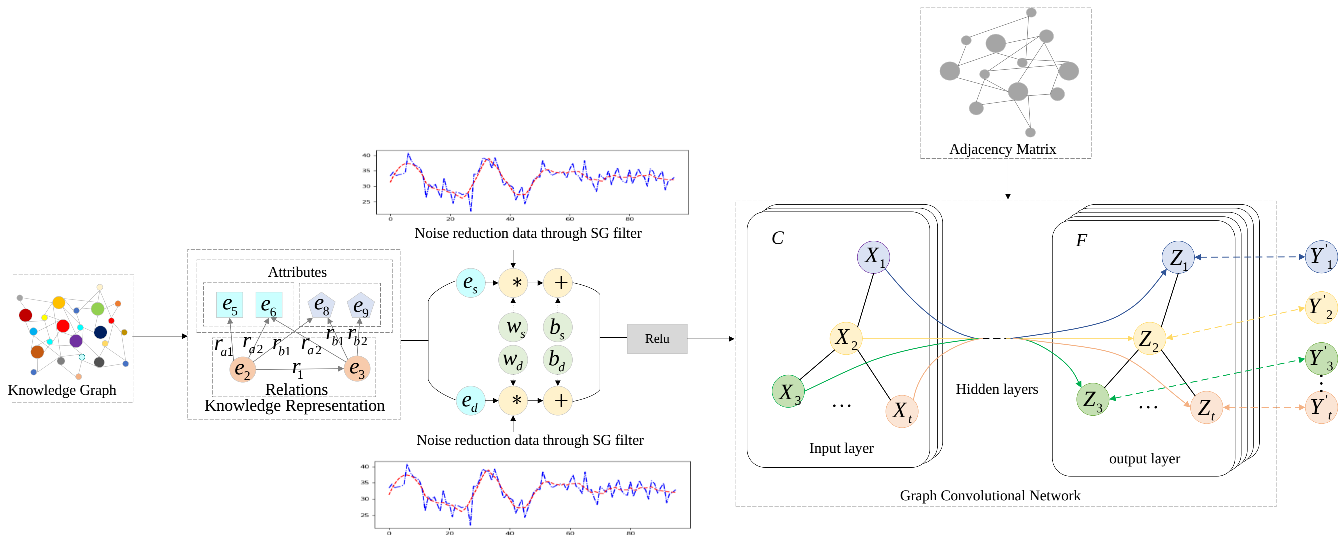


Fig. 11. Structure diagram of KGCN

researchers, which makes the LSTM model diffusely used in various domains. The LSTM model ameliorate RNN network by introducing gating mechanism. It solves the defects of RNN effectively and is a kind of RNN network with special structure. Consequently, this paper employs it to acquire the time reliance of traffic stream. Fig. 12 illustrates the LSTM structure diagram.

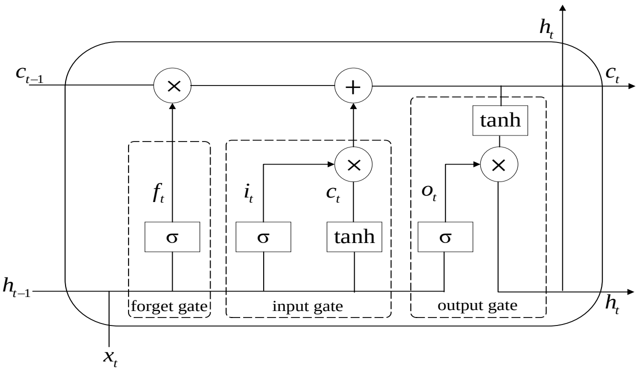


Fig. 12. LSTM structure diagram

LSTM hidden element calculation formula is as follows:

$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \tag{14}$$

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{15}$$

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{16}$$

$$\tilde{C}_t = \tanh\left(W_c\left[h_{t-1}, x_t\right] + b_c\right) \tag{17}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{18}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{19}$$

Where: i_t , f_t , o_t and $\tilde{C}_t$ indicate respectively input gate, forget gate, output gate, and temporary cell state. W_i , W_f , W_o and W_c respectively represent the weight matrix between the input vector and above three gates, and the temporary cell state. b_i , b_f , b_o and b_c represent the bias vector. C_t represents the cell's current state, h_t represents the hidden layer state at the current moment, and * represents the Hadamard product of the matrix.

D. Attention Mechanism

The attention mechanism can enhance the model's precision by adjusting the weight. With the augmenting of time series, the memory ability of the LSTM model will decline. The attention mechanism can solve the problem of memory decline by accurately locating the relationship between the local and the whole before each prediction by the model. Especially in the domain of time series forecast, attention mechanism has been widely used. Therefore, the attention mechanism has been applied by some researchers in various learning scenarios. Calculating the probability distribution of attention highlights the influence of a certain key input on output [43-44]. There are many approaches to realize the attention mechanism. Considering that the self-attention mechanism is more effective in capturing internal relationships among data or features, this paper introduced it to catch the connection between different input features.

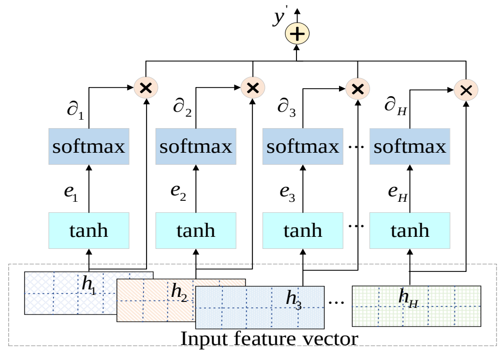


Fig. 13. Structure diagram of a self-attention mechanism

As shown in Fig. 13, for H input feature vectors h_t , t=1, 2, 3, ⋯, H, calculate the probability distribution of attention grounded on the input vector h_t at time t by the tanh function:

$$e_t = \tanh\left(w_e h_t + b_e\right) \tag{20}$$

Where: e_t represents the probability distribution value decided by the input vector h_t of the self-attention mechanism at time t, w_e and b_e are weight and bias, and e_t is normalized by softmax function. The weight coefficient corresponding to the input vector can be obtained:

$$\partial_t = \frac{\exp\left(e_t\right)}{\sum_{i=1}^{H}\exp\left(e_i\right)} \tag{21}$$

Where: ∂_t is the weight coefficient at time t. The method of weighted mean was used to fuse ∂_t and the input feature vector to get the output result $\hat{y}$:

$$\hat{y} = \sum_{t=1}^{H}\partial_t h_t \tag{22}$$

E. SGA-KGCN-LSTM Model

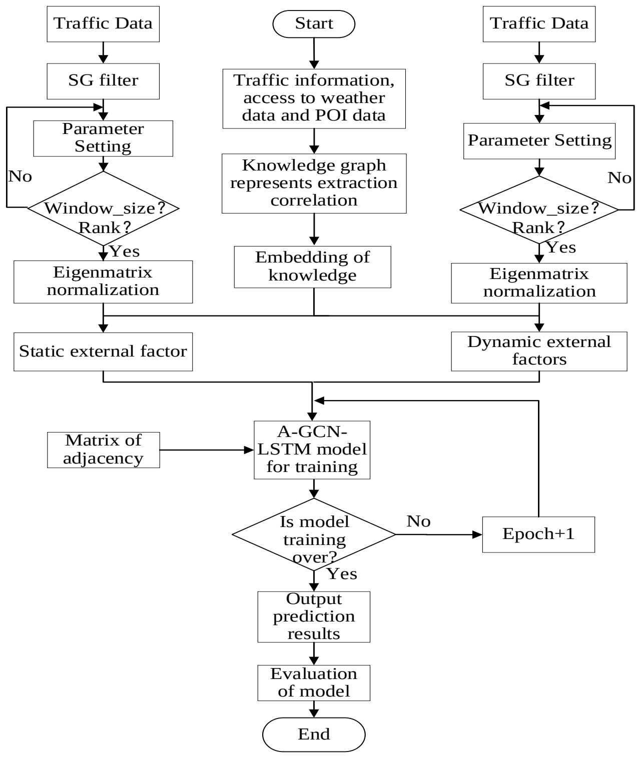


Fig. 14. Workflow of SGA-KGCN-LSTM model

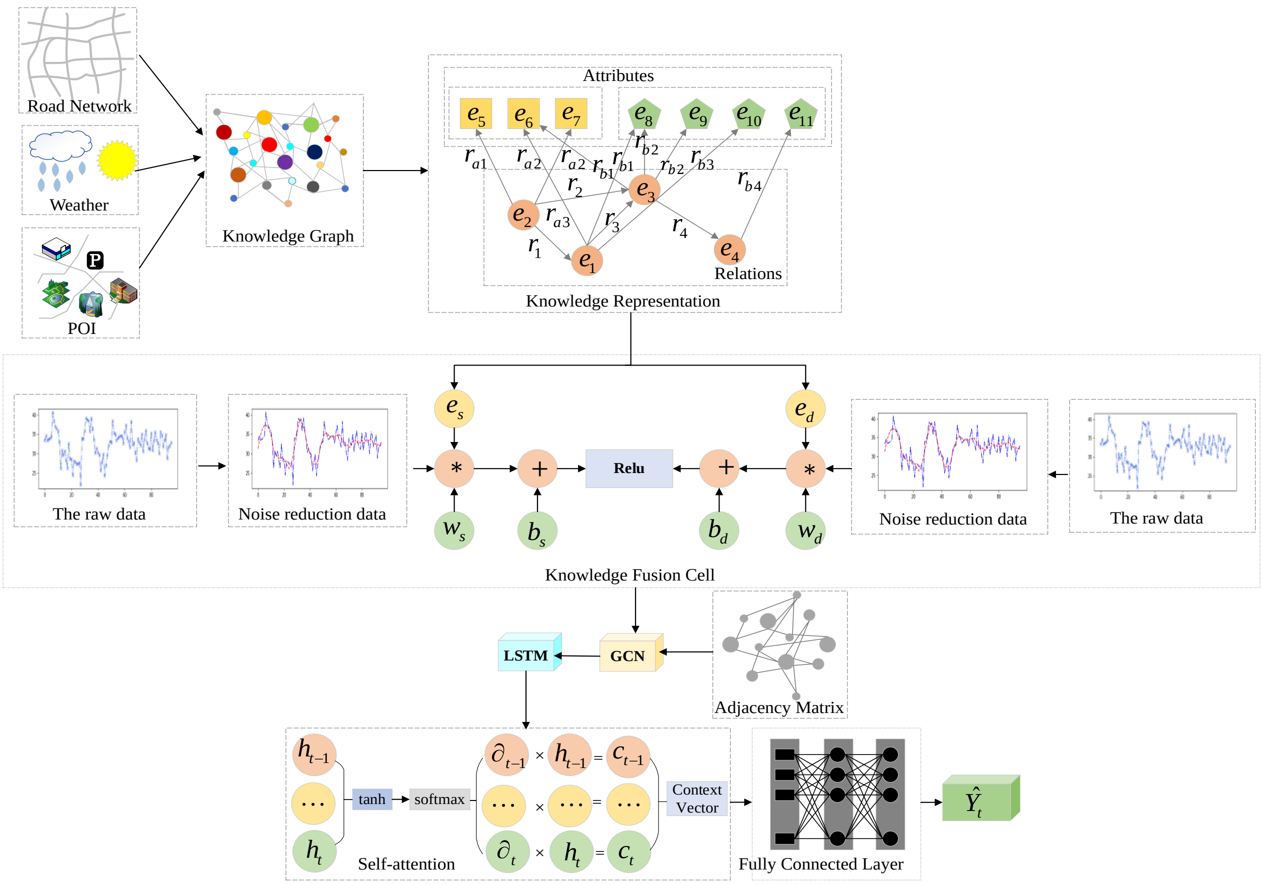


Fig. 15. Structure of SGA-KGCN-LSTM model

Drawing upon the above analysis of different models and their respective strengths, this paper proposes a combined forecasting model called SGA-KGCN-LSTM.

Existing studies have not fully considered the correlation between traffic information and external attributes. To address this issue, this paper introduces KG and obtains thee associated knowledge embedding using knowledge representation. Additionally, SG filtering is applied to decrease the unpitched sound in the initial data. To fuse the embedded knowledge with the traffic flow features after noise reduction, KF-Cell is utilized to obtain new section features. The adjacency matrix and new section features are then used as inputs for the trunk model. The traffic flow's spatiotemporal features are derived through the use of GCN and LSTM. Furthermore, to take into account the impact of various input features on model prediction, this paper incorporates the self-attention mechanism theory to acquire the interdependence among diverse characteristics. The probability distribution approach is adopted to assign appropriate weight to the input feature information. Finally, the prediction is gained by means of the fully connected layer. The workflow and structure of the SGA-KGCN-LSTM model are depicted in Fig. 14 and Fig. 15.

IV. Simulation Experiment

A. *Source of Data*

1) Road network and traffic flow data: Shenzhen Taxi Track Data set contains Shenzhen taxi track data for every street from January 1st to January 31st, 2015. The 156 main road sections in Luohu District are taken as research areas, whose connectivity is modeled by 156*156 adjacent matrix. The traffic speed time series of the selected section is formed into a feature matrix, rows represent the section index, and columns represent the time stamp index.

2) Weather data: Shenzhen weather data set, air temperature, weather condition, wind velocity, humidness, visibility, and other meteorological data of the studied area are treated as auxiliary data, climbing the weather conditions in Luohu District in January 2015. These weather conditions are recorded at 15-minute intervals and zoned into five types: sunny, overcast, misty, drizzle, and downpour.

3) POI data: Shenzhen POI data set contains POI information around each selected road section in Luohu District. The information is divided into nine types of services: food, businesses, shopping, transportation, education, life, healthcare, accommodation, and others.

B. *Construction of KG and Noise Reduction of Traffic Flow Data*

1) Construction of KG

This paper uses relational triples R, attribute triples S, and co-occurrence relation A to construct the urban KG. The relational triplet R indicates whether there is an adjacency relationship between two sections. The attribute triplet S indicates the corresponding relationship between roads and attributes. The co-occurrence relation A represents the probability that two different attributes co-occur.

Firstly, the relational triplet is constructed using whether there is an adjacency relationship between two sections. If there is an adjacency relationship between two sections, it is represented by 1; if there is no adjacency relationship, it is represented by 0. TABLE I shows the data structure of relational triples.

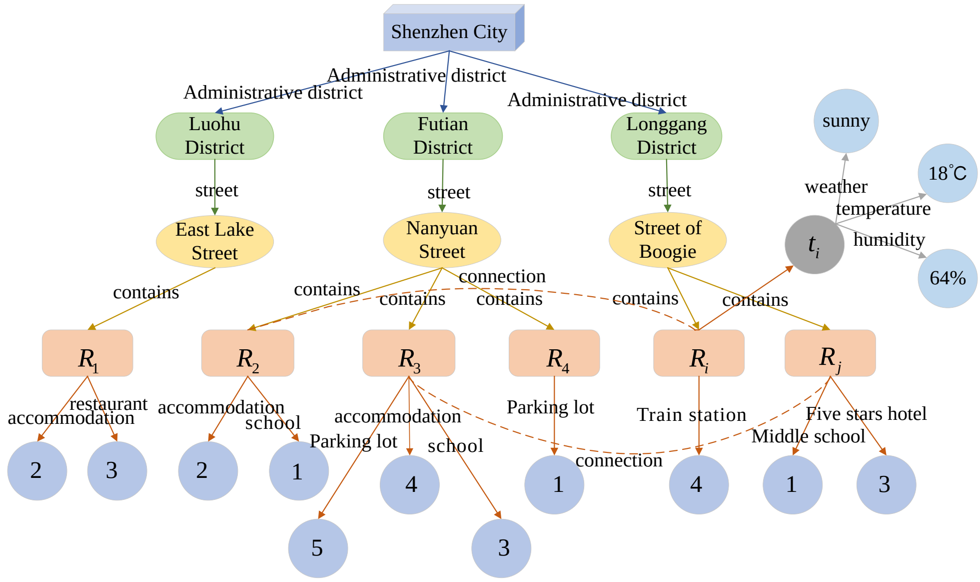


Fig. 16. Knowledge map of some cities in Shenzhen

TABLE I
THE DATA STRUCTURE OF RELATIONAL TRIPLES

adj	90217	90218	95968	95969
90217	1	0	1	1
90218	0	1	1	0
95968	1	0	1	1
95969	1	0	1	1

Secondly, a property triplet is constructed with the number of sections, categories, and poi. For example, S_1= (v_3, parking, 5), S_2= (v_3, accommodation, 4), and S_3= (v_3, schools, 3) indicate that there are five parking lots, four accommodations, and three schools near section v_3. In addition, (v_i, weather conditions, time) and (t_i, weather, sunny) are also used to construct KG. TABLE II shows the data structure of the road 92231 attribute triplet.

TABLE II
TRIPLES OF ROAD PROPERTIES

Head Entity	Attribute	Attribute Value
92231	life services	3
92231	shopping services	4
92231	accommodation services	1
92231	catering services	3
92231	transportation services	5

Finally, the co-occurrence relationship of different attributes is constructed. For example, in v_3, A = (parking lot, school, p) means that the probability of parking lot and school co-occurring in v_3 is p. TABLE III shows the co-occurrence relationship of two different attributes. In TABLE III, A = (shopping services, accommodation services, 0.004) means that the probability of shopping services and accommodation services co-occurring is 0.004.

TABLE III
CO-OCCURRENCE RELATIONSHIP OF TWO DIFFERENT ATTRIBUTES

Attribute	Attribute	Probability of co-occurrence
life services	shopping services	0.012
shopping services	accommodation services	0.004
life services	catering services	0.009
shopping services	transportation services	0.020
catering services	transportation services	0.015

By constructing road relation triples, road attribute triples, and co-occurrence relations between two different attributes, the KG of some cities in Shenzhen is finally obtained, as shown in Fig. 16.

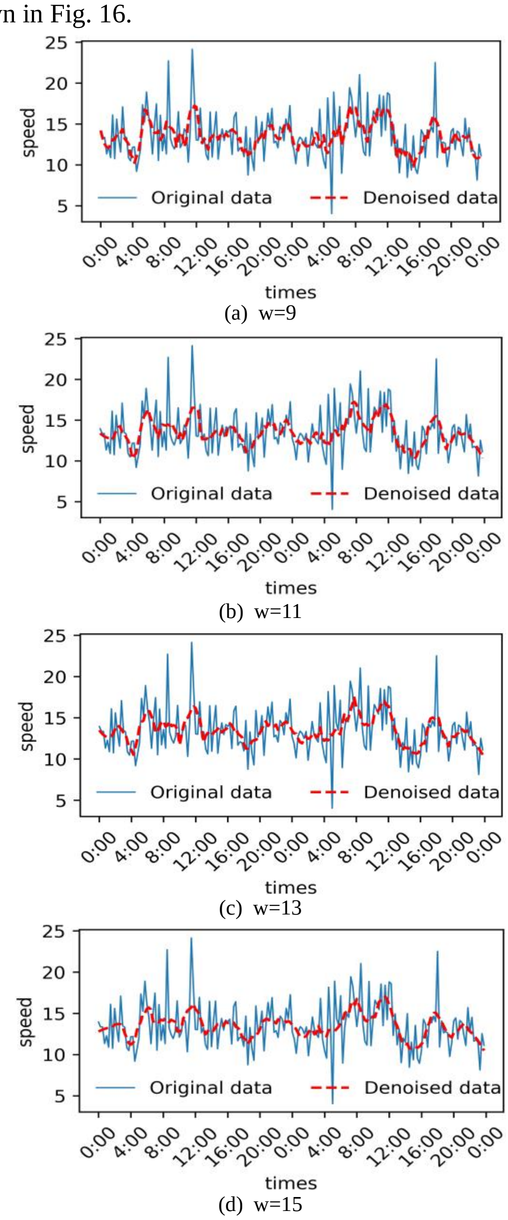


Fig. 17. Influence of window length on noise reduction

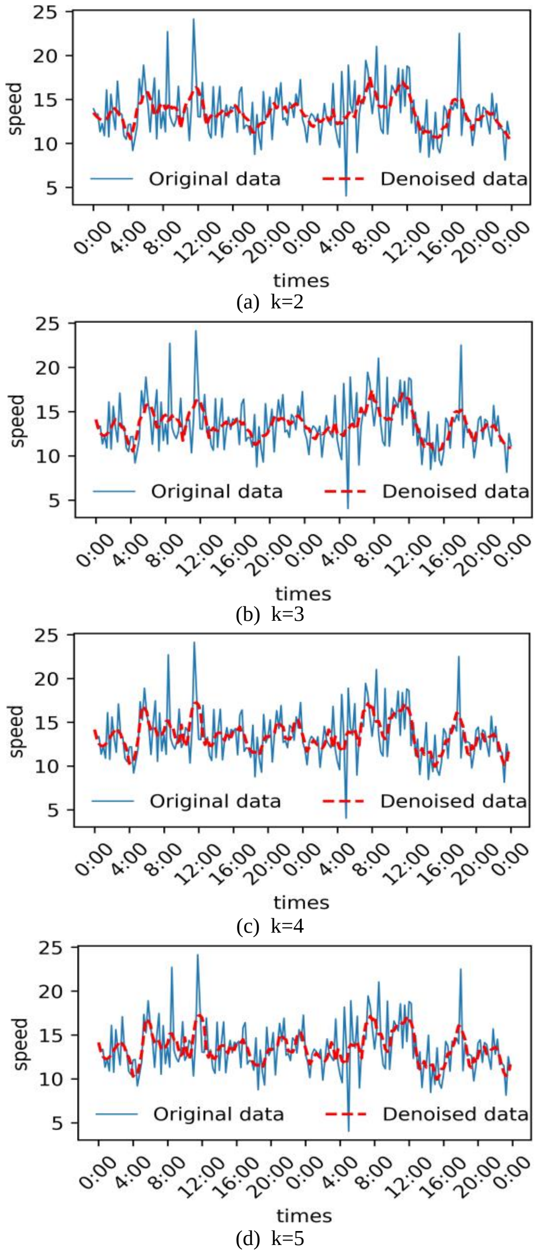
The constructed KG is embedded into knowledge.

Considering the diversity of external factors, the embedded information is divided into static and dynamic external factors and then integrated with traffic flow characteristics. Since the acquired traffic flow data will be polluted by noise, we should de-noise the traffic flow data before fusion.

2) Noise reduction of traffic flow data

Excessive noise will lead to sudden data jitter, affecting traffic flow prediction. This article utilizes the SG filter for reducing the yawp in the original data.

The effect of window length on noise reduction is shown in Fig. 17. Two days' traffic data of a road were selected, and noise reduction was performed by SG filter. The parameter w represents the window length of the filter. When w is smaller, the smooth curve is closer to the actual curve. When the w is larger, the smooth effect will be more powerful. It can be seen when w is 9 and 11, the smoothing effect is poor, and when w is 15, the curve is too smooth, so the value of w is determined to be 13.



(a) k=2
(b) k=3
(c) k=4
(d) k=5

Fig. 18. Influence of polynomial order on noise reduction

Parameter k is the k-order polynomial fitting for the data points in the window. When k gets bigger, the curve is closer to the practical curve. When the value of k is smaller, the curve is smoother. When the value of *k* is large, the fitting problem will occur due to the limitation of the window length, thus affecting the model performance.

The influence of polynomial order on noise reduction is shown in Fig. 18. In this paper, and order *k* is selected from the set {2,3,4,5}. It can be seen when *k* is 4 and 5, the smoothing effect of the curve is poor. When *k* is 2 and 3, the effect is not different, and the experiment determines the specific value.

C. Model Evaluation and Experimental Setup

1) Index of evaluation

To assess the predictive precision of the SGA-KGCN-LSTM model, this paper employs four evaluation indicators: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Accuracy, and the determination coefficient R^2. The specific definitions are as follows:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2} \tag{23}$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{24}$$

$$Accuracy=1-\frac{\left\|y-\hat{y}\right\|}{\left\|y\right\|F} \tag{25}$$

$$R^2=1-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}_i\right)^2} \tag{26}$$

Where: y_i denotes the real value, $\hat{y}_i$ denotes the output value, $\bar{y}_i$ denotes the mean value, n is the number of selected samples, and ||• ||F represents the Frobenius norm.

2) Parameter setting

In this paper, taxi trajectory data from the Luohu District of Shenzhen was selected, and the training set consisted of 80% of the dataset, with the remaining 20% allocated to the test set. SG filter is utilized to decrease the noise of primordial data. Through analyzing the effect of window length on traffic stream de-noising, the window length w of the filter is set to 13. Similarly, based on the effect of polynomial fitting order on traffic stream denoising, the polynomial fitting order k is determined to be 2. Set the epoch to 1500. Finally, using experimental search, select the values for learning rate, hidden unites, and batch size. After multiple experiments, the learning rate is determined to be 0.001, the hidden units are 32, and the batch size is 32.

3) Analysis of parameters

Since the learning rate, the quantity of hidden units, and batch size significantly influence the model's performance during the experiment, their values are analyzed using experimental search.

The influence of the learning rate on the predicted precision is displayed in Fig. 19. Test the training ability of the model in the set {0.0001, 0.001, 0.005, 0.01}, and analyze the precision changes of the model. It can be seen when the learning rate is 0.001, the values of the evaluation indicator RMSE and MAE are the minima, and the values of Accuracy and R^2 are the maximum. At this moment, the SGA-KGCN-LSTM model has the best predictive performance.

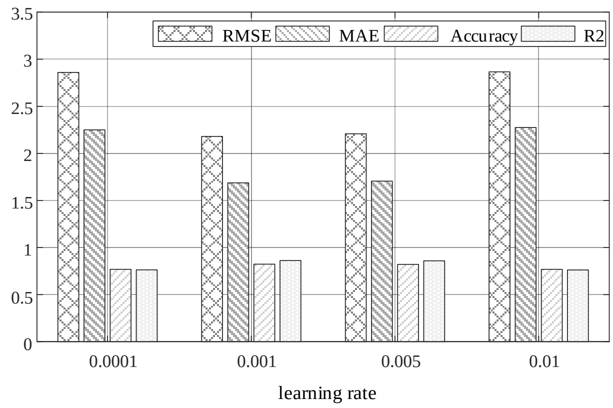


Fig. 19. Influence of learning rate on predicted performance

Fig. 20 illustrates how the predicted performance is influenced by the quantity of hidden units when learning rate is 0.001. Test the drilling ability of SGA-KGCN-LSTM of hidden units in the set {8,16,32,64}. We can find that the model's performance has little difference when hidden units are 32 and 64. However, when units are 32, the value of MAE is 0.003 less than that when units are 64, so the number of hidden units is determined to be 32.

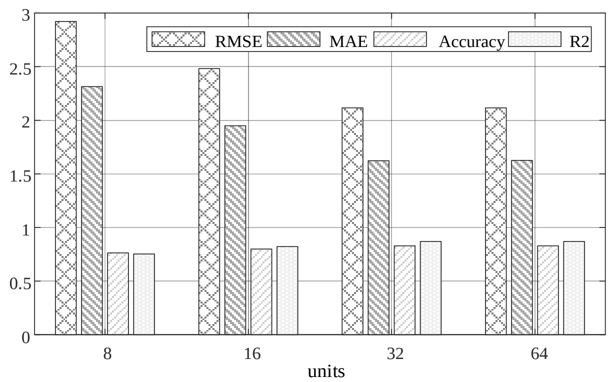


Fig. 20. Influence of hidden unit on predicted performance

When the learning rate and hidden units are 0.001 and 32, severally, the impaction of batch size on the predicted performance is shown in Fig. 21. Analyze the evaluation metrics of the SGA-KGCN-LSTM model when the batch size takes on the values in the set {8, 16, 32, 64}. We can find the precision of the SGA-KGCN-LSTM is the best when the batch size is 32, so it determined to be 32.

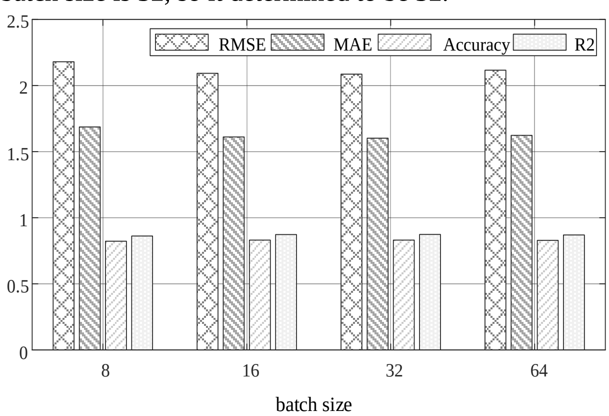


Fig. 21. Impact of batch size on prediction performance

D. *Experimental Analysis*

1) Experimental results

Draw the RMSE of the SGA-KGCN-LSTM model and the loss rate and accuracy of the test set.

Fig. 22 reveals the performance of the SGA-KGCN-LSTM model. It is observed that from (a) and (b) that RMSE and test-loss rapidly decrease within 400 batches and converge when the epoch is 1500. As seen from (c), when the epoch is 500, the test accuracy increases rapidly, reaching an accuracy of about 83.1% after 1500 training sessions.

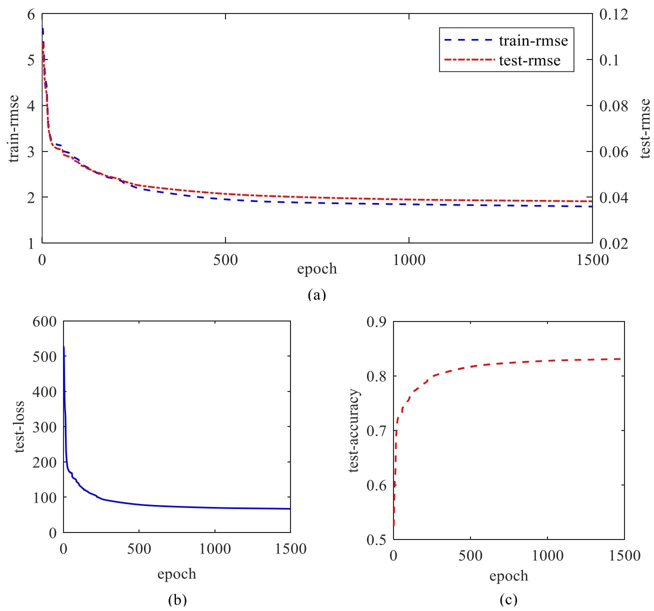


Fig. 22. Performance of the SGA-KGCN-LSTM model

2) Comparative analysis of baseline models

To demonstrate the predictive precision of the SGA-KGCN-LSTM model, seven benchmark models are selected for comparative analysis with the model presented in this paper. The seven benchmark models are described as follows:

- HA: Historical Average Model
- SVR: Support Vector Regression
- LSTM: Long Short-Term Memory Network
- GCN: Graph Convolutional Network
- GRU: Gated Recurrent Unit
- T-GCN [24]: Temporal Graph Convolutional Network
- KST-GCN [42]: Knowledge-Driven Spatial-Temporal Graph Convolutional Network

It is observed that TABLE IV that RMSE and MAE are reduced by about 2.148 and 1.178 in model SGA-KGCN-LSTM model compared with model HA. Compared with the SVR model, Accuracy and R^2 were improved by about 12% and 0.031. In comparison to the LSTM model, RMSE and MAE are reduced by about 5.598 and 4.556. Compared with the GCN model, Accuracy and R^2 are increased by 25.4% and 0.211. In comparison to the GRU model, RMSE and MAE decreased by about 4.45 and 3.255. Compared with the T-GCN model, Accuracy and R^2 are increased by about 11.9% and 0.031. In comparison to the KST-GCN model, the SGA-KGCN-LSTM model shows a reduction of approximately 1.982 and 1.274 in terms of RMSE and MAE, respectively, while Accuracy and R^2 are increased by about 10.9% and 0.047. In comparison to the above reference models, the SGA-KGCN-LSTM model has the best evaluation indexes and predictive precision.

TABLE IV
PREDICTION PERFORMANCE OF DIFFERENT MODELS

Model of prediction	RMSE	MAE	Accuracy	R^2
HA	4.234	2.780	0.705	0.836
SVR	4.140	2.688	0.711	0.843
LSTM	7.306	5.833	0.508	0.508
GCN	6.066	4.524	0.577	0.663
GRU	6.536	4.857	0.545	0.609
T-GCN	4.140	2.815	0.712	0.843
KST-GCN	4.068	2.876	0.722	0.827
SGA-KGCN-LSTM	**2.086**	**1.602**	**0.831**	**0.874**

The visualization analysis of the SGA-KGCN-LSTM model and the above reference model is carried out. As shown in Fig. 23, it is observed intuitively that the SGA-KGCN-LSTM model makes better predictions.

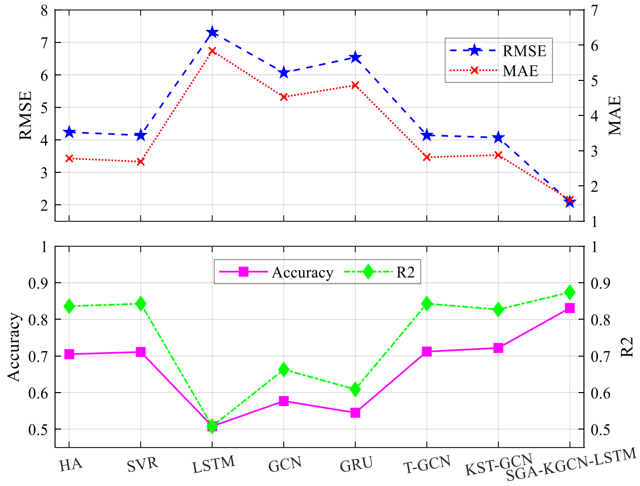


Fig. 23. Comparative analysis of errors of different models

3) Comparative analysis of ablation experiments

a. The effect of KG on the model

To verify the impaction of KG on the SGA-KGCN-LSTM model, a comparison of errors between the SGA-KGCN-LSTM model and the SGA-GCN-LSTM model (without KG) is conducted.

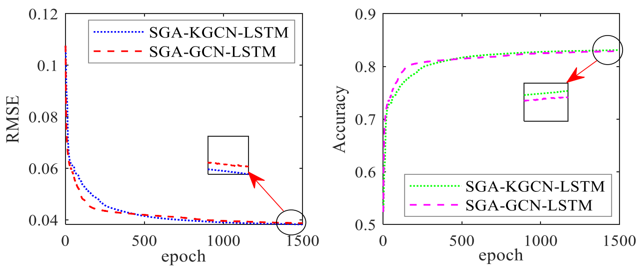


Fig. 24. Comparison of errors of KG to the model

The error comparison of KG to the model is shown in Fig. 24. We can find that RMSE and Accuracy indexes of the SGA-KGCN-LSTM model are better than the SGA-GCN-LSTM model with the integration of full knowledge because the SGA-KGCN-LSTM model considers the impact of the relationship between traffic data and outside factors on traffic stream forecasting.

b. The impact of the self-attention mechanism on the model

To verify the self-attention mechanism's influence on the SGA-KGCN-LSTM model, a comparison of errors between the SGA-KGCN-LSTM model and the SG-KGCN-LSTM model (without the self-attention mechanism) is conducted under the condition of full knowledge.

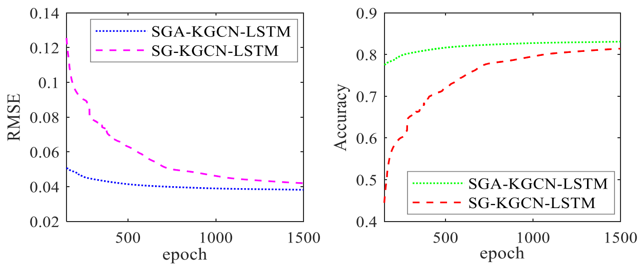


Fig. 25. Comparison of errors of self-attention mechanism to the model

When the epoch is smaller than 150, the RMSE drops too fast. Therefore, only the analysis is performed when the epoch exceeds 150. As depicted in Fig.25, both RMSE and Accuracy indexes of the SGA-KGCN-LSTM are better than those of the SG-KGCN-LSTM model without the introduction of self-attention mechanism.

c. SG effect on the model

To verify the impact of the SG filter on the SGA-KGCN-LSTM model, an error comparison between the A-KGCN-LSTM and the SGA-KGCN-LSTM model without the SG filter is carried out under the condition of full knowledge.

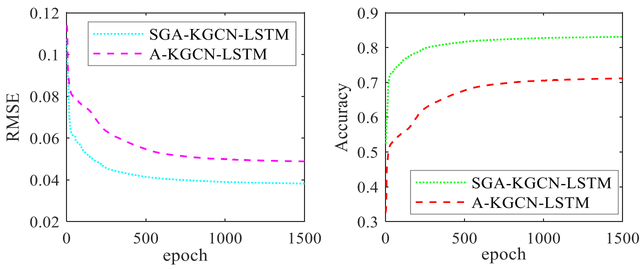


Fig. 26. Error comparison of SG to the model

The error comparison between SG and the model is displayed in Fig. 26. It is detectable that the SGA-KGCN-LSTM model incorporated exhibits superior RMSE and Accuracy compared to the A-KGCN-LSTM model without the SG filter incorporated.

d. The influence of ablation models incorporating different knowledge on the model

In order to verify the predictive precision of diverse models when integrating different types of knowledge in ablation experiments, this paper further conducts comparative analysis by integrating KGCN-LSTM, A-KGCN-LSTM, SG-KGCN-LSTM, and SGA-KGCN-LSTM models with different types of knowledge, and analyzing their respective predictive performances.

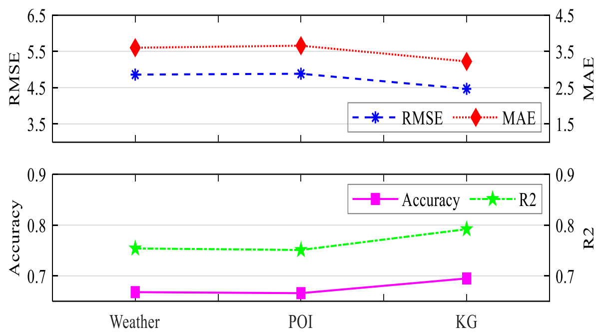


Fig. 27. Comparison of KGCN-LSTM model errors

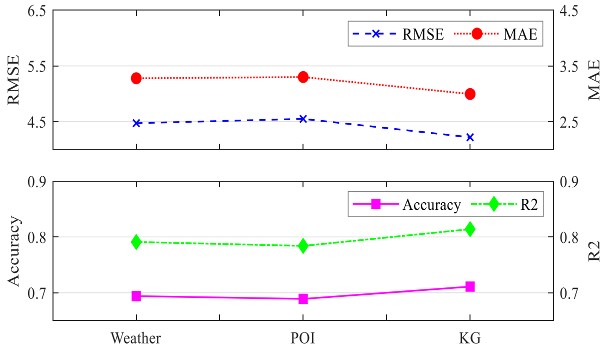


Fig. 28. A-KGCN-LSTM model error comparison

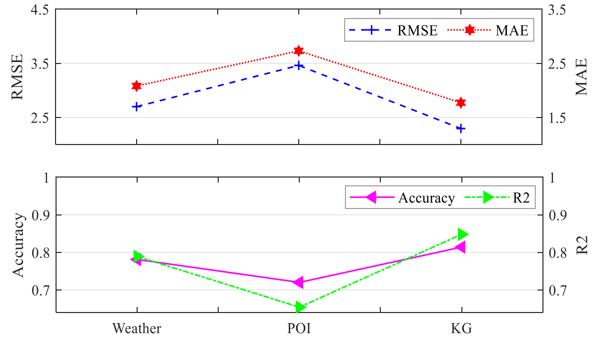


Fig. 29. Comparative of SG-KGCN-LSTM model errors

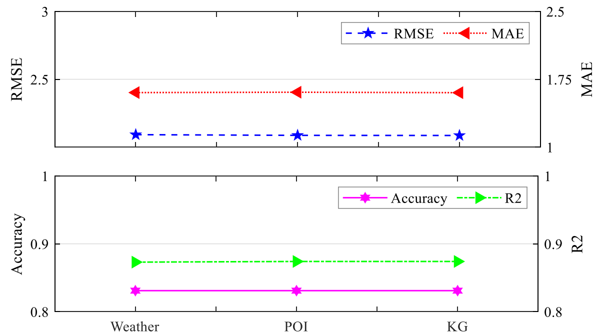


Fig. 30. Comparative of SGA-KGCN-LSTM model errors

When different ablation models incorporate different knowledge, the influence on the models is shown in Fig. 27, Fig. 28, Fig. 29, and Fig. 30. The KGCN-LSTM model without self-attention mechanism and SG filter, A-KGCN-LSTM model without SG filter, SG-KGCN-LSTM model without self-attention mechanism and SGA-KGCN-LSTM model are embedded with different knowledge, and analyze the impact of each case on traffic volume forecast. It is observed that the evaluation indicators of each model with the integration of full knowledge are better than other cases.

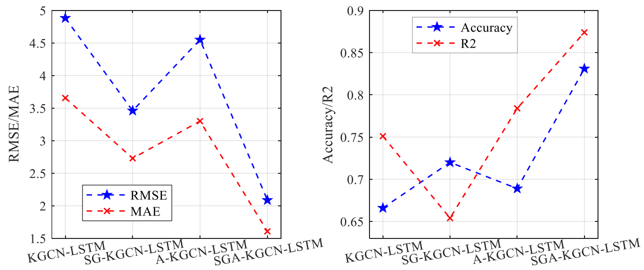


Fig. 31. Error comparison of different models under Poi

To visually see the error analysis of evaluation indexes of different ablation models in the same situation, the traffic characteristics integrated with Poi, weather, and full knowledge are input into different models for comparative analysis.

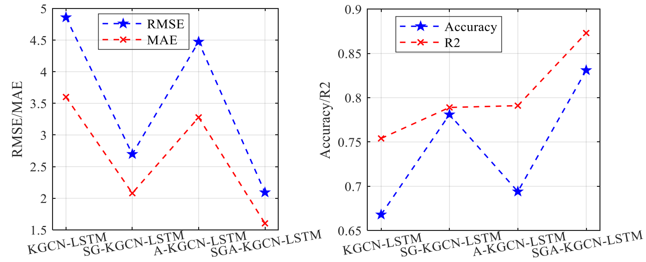


Fig. 32. Comparison of errors of different models under weather conditions

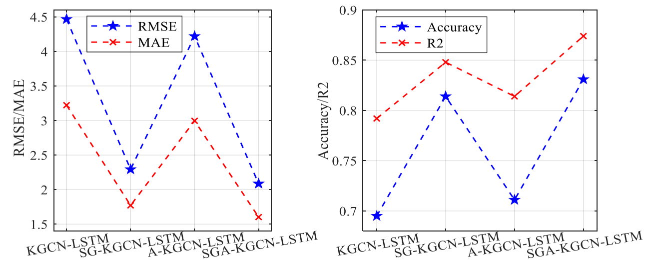


Fig. 33. Error comparison of different models in the case of full knowledge

As shown in Fig. 31, Fig. 32, and Fig. 33, when the KGCN-LSTM model, SG-KGCN-LSTM model, A-KGCN-LSTM model, and SGA-KGCN-LSTM model are integrated into POI, weather, and full knowledge respectively, the performance of SGA-KGCN-LSTM model is the best. This model's RMSE, MAE, Accuracy, and R2 are superior to others. Through the above analysis, this paper finally takes the traffic flow characteristics integrated with the full knowledge as the input of this paper model. The SGA-KGCN-LSTM model is used to forecast the future traffic stream in a short time.

The above visualization analysis of ablation experiments can intuitively see the error comparison of each model's evaluation indexes RMSE, MAE, Accuracy, and R^2 when different models incorporate different knowledge and the same knowledge. However, when we want to know the specific value of the evaluation index of a particular model when different knowledge is incorporated, we cannot see from the visualization analysis of the ablation experiment. For example, it isn't easy to see the RMSE and MAE of the SG-KGCN-LSTM without the self-attention mechanism when the weather is integrated. Therefore, we present it in a tabulated form. We can see the specific values of evaluation indicators through the table when the model incorporates different knowledge.

In TABLE V, when the traffic flow features integrated with full knowledge are input into the KGCN-LSTM model, it is observed that RMSE is about 0.392 and 0.416 less than that combined with weather and POI knowledge. When the traffic features integrated with the full knowledge were input into the A-KGCN-LSTM model, MAE decreased by approximately 0.282 and 0.304 compared to the case with the combined knowledge of weather and POI. When the traffic features integrated with the full knowledge were input into the SG-KGCN-LSTM model, the accuracy was approximately 3.3% and 9.4% higher than that in the case combined with the weather and POI knowledge. When the all-knowledge traffic features were input into the

TABLE V
ERROR COMPARISON OF DIFFERENT MODELS UNDER DIFFERENT CIRCUMSTANCES

Model of prediction		RMSE	MAE	Accuracy	R^2
KGCN-LSTM	POI	4.883	3.656	0.666	0.751
	Weather	4.859	3.600	0.668	0.754
	Kg	4.467	3.222	0.695	0.792
A-KGCN-LSTM	POI	4.550	3.300	0.689	0.784
	Weather	4.473	3.278	0.694	0.791
	Kg	4.220	2.996	0.711	0.814
SG-KGCN-LSTM	POI	3.460	2.729	0.720	0.654
	Weather	2.698	2.080	0.781	0.789
	Kg	2.293	1.771	0.814	0.848
SGA-KGCN-LSTM	POI	2.087	1.607	0.831	0.874
	Weather	2.092	1.603	0.831	0.873
	Kg	**2.086**	**1.602**	**0.831**	**0.874**

SGA-KGCN-LSTM model, the coefficient of determination R2 improved by about 0.001 compared to the case with weather knowledge. Through the above comparative analysis, the traffic features integrated with the full knowledge are finally utilized as the input of the SGA-KGCN-LSTM model.

V. CONCLUSION

To enhancing the precision of short-term traffic volume forecast, this study suggests the SGA-GCN-LSTM model. Comparative analysis between this model and benchmark models demonstrates superior evaluation metrics. Additionally, to validate the efficacy of the suggested method, an ablation experiment is conducted. The characteristics of the model are as follows:

(1) KG is used to realize the fusion of correlation between traffic information and external attributes and traffic features. In contrast with the SGA-GCN-LSTM model without KG, RMSE and Accuracy indexes of SGA-KGCN-LSTM are minimum.

(2) The SGA-KGCN-LSTM model applies SG filter to reduce the unpitched sound in the original data. Compared to the A-KGCN-LSTM model without SG filter, the SGA-KGCN-LSTM model exhibits better performance in terms of RMSE and Accuracy. These results suggest that the SG filter enhances the predictive precision of the SGA-KGCN-LSTM model.

(3) By combining the strengths of GCN and LSTM models, the SGA-KGCN-LSTM model proficiently captures the spatial-temporal characteristics of traffic stream while introducing self-attention mechanism to extract relationships between different input characteristics. Comparative analysis between the SGA-KGCN-LSTM and SG-KGCN-LSTM models reveals that the SGA-KGCN-LSTM model outperforms the SG-KGCN-LSTM model in terms of RMSE and Accuracy.

(4) The SGA-KGCN-LSTM model is contrasted with the reference model, and the ablation experiment is also compared. From the perspective of the knowledge graph, self-attention mechanism, and SG filter, as well as the situation of different models with different knowledge, it is concluded that each evaluation index of the SGA-KGCN-LSTM model is better than other methods.

Although the SGA-KGCN-LSTM model takes about the effect of weather and POI factors on traffic flow prediction, it does not take about traffic accidents and epidemic situations, etc. In the future traffic flow forecasting research, more external factors will be considered, sequentially further enhance the precision of short-term traffic volume forecast.

REFERENCES

[1] W. Du, Q. Zhang, Y. Chen, Z. Ye, "An urban short-term traffic flow prediction model based on wavelet neural network with improved whale optimization algorithm," *Sustainable Cities and Society*, vol. 69, pp. 102858, (2021)

[2] L. Han, Y.-S. Huang, "Short-term traffic flow prediction of road network based on deep learning," *IET Intelligent Transport Systems*, vol. 14, pp. 495-503, (2020)

[3] W. Wu, X. Fan, Y. Xue, Y. Huang, "An attention mechanism-based method for predicting traffic flow by GCN," *2021 40th Chinese Control Conference (CCC)*, pp. 8410-8415

[4] H. Zhao, D. Zhan, C. Shi, "Review of short-term traffic flow prediction model," *Urban Rapid Rail Transit*, vol. 32, pp. 50-54, (2019)

[5] X. Luo, D. Li, S. Zhang, "Traffic Flow Prediction during the Holidays Based on DFT and SVR," *Journal of Sensors*, vol. 2019, pp. 6461450, (2019)

[6] S. Shahriari, M. Ghasri, S.A. Sisson, T.H. Rashidi, "Ensemble of ARIMA: combining parametric and bootstrapping technique for traffic flow prediction," *Transportmetrica A: Transport Science*, vol. 16, pp. 1552 - 1573, (2020)

[7] Z. Mingheng, Z. Yaobao, H. Ganglong, C. Gang, "Accurate Multisteps Traffic Flow Prediction Based on SVM," *Mathematical Problems in Engineering*, vol. 2013, pp. 418303, (2013)

[8] H. Lu, Y. Han, K. Zhang, L. Zhang, Y. Ding, "Short-term Traffic Flow Prediction Based on VMD-BiLSTM-BLS Model," *Computer Systems and Applications*, vol. 31, pp. 238-245, (2022)

[9] W. Zhang, Y. Yu, Y. Qi, F. Shu, Y. Wang, "Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning," *Transportmetrica A Transport Science*, vol. 15, pp. 1688-1711, (2019)

[10] Y. Tian, K. Zhang, J. Li, X. Lin, B. Yang, "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, vol. 318, pp. 297-305, (2018)

[11] L. Mou, P. Zhao, H. Xie, Y. Chen, "T-LSTM: A Long Short-Term Memory Neural Network Enhanced by Temporal Information for Traffic Flow Prediction," *IEEE Access*, vol. 7, pp. 98053-98060, (2019)

[12] J. Zheng, M. Huang, "Traffic Flow Forecast Through Time Series Analysis Based on Deep Learning," *IEEE Access*, vol. 8, pp. 82562-82570, (2020)

[13] W. Wei, H. Wu, H. Ma, "An AutoEncoder and LSTM-Based Traffic Flow Prediction Method," *Sensors*, vol. 19, pp. 2946, (2019)

[14] J. Zhang, S. Qu, Z. Zhang, S. Cheng, "Improved genetic algorithm optimized LSTM model and its application in short-term traffic flow prediction," *PeerJ Computer Science*, vol. 8, pp. 1048, (2022)

[15] X. Chen, H. Chen, Y. Yang, H. Wu, W. Zhang, J. Zhao, Y. Xiong, "Traffic flow prediction by an ensemble framework with data denoising and deep learning model," *Physica A: Statistical Mechanics and its Applications*, vol. 565, pp. 125574, (2021)

[16] X. Chen, S. Wu, C. Shi, Y. Huang, Y. Yang, R. Ke, J. Zhao, "Sensing Data Supported Traffic Flow Prediction via Denoising Schemes and ANN: A Comparison," *IEEE Sensors Journal*, vol. 20, pp. 14317-14328, (2020)

[17] J. Bi, H. Yuan, K. Xu, H. Ma, M. Zhou, "Large-scale Network Traffic Prediction With LSTM and Temporal Convolutional Networks," *2022 International Conference on Robotics and Automation (ICRA)*, pp. 3865-3870

[18] Q. Dong, Y. Lin, J. Bi, H. Yuan, "An Integrated Deep Neural Network Approach for Large-Scale Water Quality Time Series Prediction," *2019 IEEE International Conference on Systems, Man and Cybernetics* (SMC), pp. 3537-3542

[19] X. Luo, D. Li, Y. Yang, S. Zhang, "Spatiotemporal Traffic Flow Prediction with KNN and LSTM," *Journal of Advanced Transportation*, pp. 4145353, (2019)

[20] G. Dai, C. Ma, X. Xu, "Short-Term Traffic Flow Prediction Method for Urban Road Sections Based on Space-Time Analysis and GRU," *IEEE Access*, vol. 7, pp. 143025-143035, (2019)

[21] M. Lv, Z. Hong, L. Chen, T. Chen, T. Zhu, S. Ji, "Temporal Multi-Graph Convolutional Network for Traffic Flow Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, pp. 3337-3348, (2021)

[22] W. Zhuang, Y. Cao, "Short-Term Traffic Flow Prediction Based on CNN-BILSTM with Multicomponent Information," *Applied Sciences*, vol. 12,

[23] C. Chai, C. Ren, C. Yin, H. Xu, Q. Meng, J. Teng, G. Gao, "A Multifeature Fusion Short-Term Traffic Flow Prediction Model Based on Deep Learnings," *Journal of Advanced Transportation*, vol. 2022, (2022)

[24] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, H. Li, "T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, pp. 3848-3858, (2020)

[25] H. Hu, Z. Lin, Q. Hu, Y. Zhang, "Attention Mechanism With Spatial-Temporal Joint Model for Traffic Flow Speed Prediction," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-10, (2021)

[26] X. Zhang, Y. Xu, Y. Shao, "Forecasting traffic flow with spatial–temporal convolutional graph attention networks," *Neural Computing and Applications*, vol. 34, pp. 15457-15479, (2022)

[27] H. Zheng, F. Lin, X. Feng, Y. Chen, "A Hybrid Deep Learning Model With Attention-Based Conv-LSTM Networks for Short-Term Traffic Flow Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, pp. 6910-6920, (2021)

[28] L.N.N. Do, H.L. Vu, B.Q. Vo, Z. Liu, D. Phung, "An effective spatial-temporal attention based neural network for traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 108, pp. 12-28, (2019)

[29] J. Yu, H. Wei, H. Guo, Y. Cai, "Urban Traffic State Prediction Based on SA-LSTM," *IOP Conference Series: Earth and Environmental Science*, vol. 783, pp. 012153, (2021)

[30] W. Zhang, P. Zhang, Y. Yu, X. Li, S.A. Biancardo, J. Zhang, "Missing Data Repairs for Traffic Flow With Self-Attention Generative Adversarial Imputation Net," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 7919-7930, (2022)

[31] S. Guo, Y. Lin, N. Feng, C. Song, H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, pp. 114

[32] B. Vijayalakshmi, K. Ramar, N.Z. Jhanjhi, S. Verma, M. Kaliappan, K. Vijayalakshmi, S. Vimal, Kavita, U. Ghosh, "An attention-based deep learning model for traffic flow prediction using spatiotemporal features towards sustainable smart city" *International Journal of Communication Systems*, vol. 34, pp. e4609, (2021)

[33] Q. Wang, K. Tian, C. Zhu, Y. Wei, "Research on Short-time Traffic Flow Prediction Based on Multi-factors," *Computer Engineering and Applications*, vol. 58, pp. 309-316, (2022)

[34] M.-A. Duhayyim, A.-A. Albraikan, F.-N. Al-Wesabi "Modeling of Artificial Intelligence Based Traffic Flow Prediction with Weather Conditions," *Computers, Materials and Continua*, vol. 71, pp. 3953--3968, (2022)

[35] Q. Wang, Y. Wu, C. Zhu, Y. Wang, "Research on Short-time Traffic Flow Prediction Based on External Attributes," *Computer Application Research*, vol. 39, pp. 2974-2978, (2022)

[36] K. Wang, C. Ma, Y. Qiao, X. Lu, W. Hao, S. Dong, "A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction," *Physica A: Statistical Mechanics and its Applications*, vol. 583, pp. 126293, (2021)

[37] J. Ye, S. Xue, A. Jiang, "Attention-based spatio-temporal graph convolutional network considering external factors for multi-step traffic flow prediction," *Digital Communications and Networks*, vol. 8, pp. 343-350, (2022)

[38] Z. Gui, Z. Li, L. Guo, "Short-term Traffic flow Prediction based on ACGRU Model," *Computer Engineering and Applications*, vol. 56, pp. 260-265, (2020)

[39] J. Zhang, X. Zhang, C. Wu, Z. Zhai, "Knowledge map building technical review," *Computer engineering*, vol. 48, pp. 23-37, (2022)

[40] Y. Lin, Z. Liu, M. Sun, "Knowledge representation learning with entities, attributes and relations," *25th International Joint Conference on Artificial Intelligence*, vol. 1, pp. 2866-2872. (2016)

[41] Y. Lin, Z. Liu, M. Sun, Y. Liu, X. Zhu, "Learning Entity and Relation Embeddings for Knowledge Graph Completion," in *AAAI Conference on Artificial Intelligence*, (2015)

[42] J. Zhu, X. Han, H. Deng, C. Tao, L. Zhao, P. Wang, T. Lin, H. Li, "KST-GCN: A Knowledge-Driven Spatial-Temporal Graph Convolutional Network for Traffic Forecasting," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-11, (2022)

[43] J. Liu, J. Liu, X. Luo, "Attention mechanism research progress in depth study," *Journal of engineering science*, vol. 43, pp. 1499-1511, (2021)

[44] V. Mnih, N. Heess, A. Graves, K. Kavukcuoglu, "Recurrent models of visual attention," presented at the Proceedings of the 27th International Conference on Neural Information Processing Systems, vol. 2, pp. 2204–2212, (2014)

Qingrong Wang was born in 1977, female, professor, master's supervisor, and teacher at the Department of Computer Science and Technology. Her main research direction is big data and the application of data mining in intelligent transportation.

Xiaohong Chen was born in 1995, female, master student. Her research interests are machine learning and intelligent transportation.

Changfeng Zhu was born in 1972, male, professor, doctoral supervisor, Deputy Dean of School of Traffic and Transportation Lanzhou Jiaotong University. His research interest covers rail transit organization, rail transit system optimization and decision-making.

Wei Chai was born in 1995, male, master student, main research direction is automatic intelligent control.