Multi-source Heterogeneous Data Fusion Model Based on FC-SAE

Hong Zhang, Kun Jiang*, Chuanqi Cheng, Jie Cao, Wenyue Zhang

School of Computer and Communication, Lanzhou University of Technology, China
zhanghong@lut.edu.cn, 1252711824@qq.com, 1255531310@qq.com, caoj@lut.edu.cn, 1003239499@qq.com

Abstract

Multi-source heterogeneous data has different degrees of data correlation or data conflict. How to fuse this data and fully mine its inherent meanings to obtain more accurate decision information is a problem that needs to be solved urgently. This paper proposes a multi-source heterogeneous data fusion model based on fully connected layers and sparse autoencoders (short for FC-SAE) to solve the above problem. This model can effectively improve the time series forecasting performance compared with the traditional time series forecasting model. The MAE value is reduced by 4.4% and the RMSE value is reduced by 3.7%. In terms of fusion strategy, the method that uses the sparse autoencoder as the fusion strategy reduces the MAE value by 1.7% and the RMSE value by 2.3% compared with the method that uses the fully connected layer as the fusion strategy.

Keywords: Multi-source heterogeneous, Data fusion, Deep learning, SAE, FC

1 Introduction

Under the background of big data, data types present a diversified trend, such as video, audio, text, etc. The data structure barriers between heterogeneous data make it difficult to fully explore the relevant characteristics between heterogeneous data. As an important data processing technology, data fusion technology can break the structural barriers between heterogeneous data and make full use of the relevant characteristics of heterogeneous data for analysis. At present, data fusion technology is widely used in many fields. For example, in the field of intelligent manufacturing, data fusion technology is used for data cleaning or denoising, integrated modeling, and multi-scale classification of massive, high-dimensional, multi-source heterogeneous noisy industrial data, for subsequent correlation analysis, performance prediction, optimization decision-making, etc.

Multi-source heterogeneous time series data fusion has great significance to improve the accuracy of time series forecasting. Existing time series forecasting methods have been able to establish suitable forecasting models for time series data and have very good results. However, these methods mainly focus on the trend and short-term correlation of the time series [1]. They can't explain the unexpected events that affect the change of the time series. Usually, the emergencies are described in text form [2], which contains contextual explanations of multiple patterns in the time series. Therefore, from the perspective of data fusion, the fusion of time series data and event text data can better solve the problem of time series forecasting in the real world. Text data is generally a kind of unstructured data, while time-series data is generally in a structured form. The heterogeneous gap in heterogeneous data makes the fusion of time-series data and text data challenging.

This paper aims to build a multi-source heterogeneous time series data fusion model based on deep learning to make up for the shortcomings of traditional time series forecasting that does not fully utilize multi-source information and improve the accuracy of time forecasting. Specifically, this paper focuses on solving the taxi demand problem. But the method is not only suitable for solving the taxi demand problem, but also for solving other multi-source heterogeneous data fusion problems. In the past, deep learning models usually used fully connected neural networks for data fusion. This fusion method can only perform a simple screening of the spliced features. Different from the past research, this paper proposes a multi-source heterogeneous data fusion model based on FC-SAE. This fusion method uses sparse autoencoder which can mine the deep information of the data. The main contributions of this article are as follows:

(1) This paper proposes a multi-source heterogeneous data fusion model based on deep learning which improves the accuracy of time series forecasting by fusing multi-source heterogeneous data.

(2) The FC-SAE-based multi-source heterogeneous data fusion model proposed in this paper breaks the structural barrier between time series data and text data and solves the problem of time series data and text data fusion.

(3) This paper proposes a FC-SAE-based multi-source heterogeneous data fusion model. This model uses sparse autoencoder as the fusion strategy which has a better fusion effect than the fully connected layer fusion strategy.

The rest of this article is organized as follows. In the next section, this paper reviews the articles related to data fusion. Section 3 presents the proposed model. Section 4 presents the model algorithm steps. The experimental results are presented in Section 5. Section 6 is the conclusion of the article. Funding at the end of the article (Section 7).

2 Related Works

Deep learning can automatically learn features and extract features. This makes deep learning have great advantages in different fields such as speech recognition, target detection, image recognition, and natural language processing [3]. The advantages of deep learning in processing different types of data make it particularly suitable for dealing with problems

*Corresponding Author: Kun Jiang; E-mail: 1252711824@qq.com
DOI: 10.53106/160792642022122307003

related to the fusion of multi-source heterogeneous data [4-5]. At present, data fusion methods based on deep learning have a wide range of applications in many fields. Zhang [6] et al proposed a data fusion method based on DBN, which effectively improves the recognition accuracy of the deterioration of the ball screw. Chen [7] et al. proposed a data fusion model based on Deep Convolutional Neural Network (DCNN) to improve the effect of fault diagnosis. Through DCNN, the data collected by multiple sensors are fused to effectively eliminate the influence of noise and obtain better results. Wu [8] et al. proposed a data fusion method based on Deep Long Short Term Memory (DLSTM). This method uses Long Short Term Memory (LSTM) neurons as the basic unit to build DLSTM to fuse multi-sensor data which can well capture the long-term dependence between data and extract the deep features of the data. It effectively improves prediction accuracy. In addition, researchers have also studied the fusion of time series data and text data. Ruiz [9] et al. analyzed the correlation between Web activities, stock prices, and stock trading volume. The results of the paper show that the introduction of Web activities can effectively optimize the trading strategy of the stock market. Pereira [10] et al. used LDA to learn the topic word model of event description text and used the topic word as the input of the shallow neural network. This model compares seven commonly used machine-learning models in the case of whether to consider emergencies. The results show that considering emergencies can significantly improve the prediction results. Li [11] et al. proposed a two-level information fusion approach. This method fuses stock return data, financial data, and social textual data to examine the effects of peer engagement on stock price synchronicity. The experimental results show that peer engagement, group diversity, experts, and epidemics can reduce stock price synchronicity. Ye [12] et al. proposed a multi-view ensemble learning method. This method includes the local fusion stage and the global fusion stage. In the local fusion stage, it extracts the raw features through the emotional element statistical method, Chi-square statistical method, and emoticon space mapping method. Then five base classifiers are constructed from these features. In the global fusion stage, it uses the accuracy-based weighted method to integrate the prediction results of the five classifiers. The experimental results show that the method has better performance in identifying the polarities of microblog posts.

Through the analysis of the above research content, it is found that the existing deep learning-based data fusion methods usually fuse homogeneous multi-sensor data or different abstract levels of the same data. In addition, it usually relies on a preset dictionary of certain topics or emotion-related words in the research work which is related to the fusion of time-series data and text data.

Inspired by the above content, this paper proposes a multi-source heterogeneous data fusion model based on the FC-SAE. This model uses Glove word embedding and CNN to automatically extract semantic features which are related to time series data in text data. It can avoid dependence on preset dictionaries. At the same time, the FC-SAE model uses FC to model time-series data and weather data. It can extract potential features of time series data and weather data. In the fusion stage, this model uses SAE to fuse the semantic features of text data, time-series features, and weather data features. This model uses the SAE encoding and decoding process to effectively mine semantic features and weather data features. SAE requires input equal to output. It can compress the semantic features and weather data features to extract valid features. These features have a greater impact on time-series data fluctuations. Then according to the association relationship, this model merges them to achieve the purpose of making full use of valuable information.

3 Model Building

In Figure 1, the multi-source heterogeneous data fusion model based on FC-SAE mainly includes three modules: text feature extraction, time series modeling module, and heterogeneous data feature fusion. The text feature extraction module mainly preprocesses the text data. It uses the GloVe model for word vectorization and uses CNN to extract features at different levels of abstraction. The time series modeling module uses multi-layer FC neural network to model nonlinear time series. The heterogeneous data feature fusion module uses the encoder of the SAE model to fuse text features, weather features and time-series features.

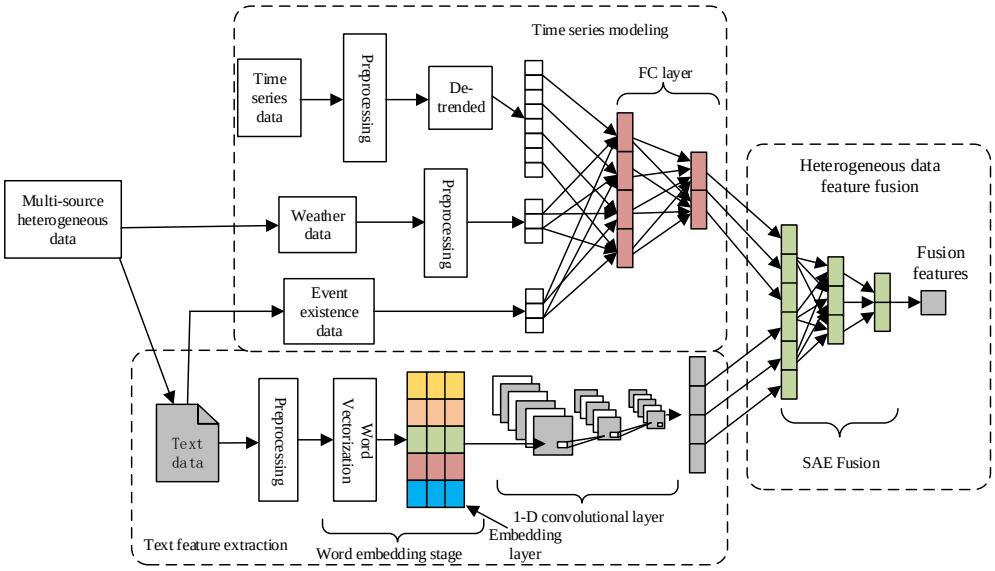

Figure 1. FC-SAE fusion model structure diagram

3.1 Text Data Feature Extraction

As one of the objects of data fusion, the feature extraction of text data is a crucial link. Text data includes useful information and noise information. The main purpose of feature extraction is to remove noise information and retain useful information. The effect of text feature extraction has a great influence on the effect of data fusion. Text data usually use word vectorization for semantic feature construction and then performs feature extraction on the constructed features. At present, there are two widely used methods for constructing semantic features: methods based on global matrix factorization and methods based on local context windows [13]. The former can use statistical information for semantic analysis, but the vocabulary category is relatively poor. The latter has good vocabulary analogy ability, but it can't make full use of global statistical information. The GloVe word embedding model combines the advantages of the above two types of methods and has a good word vectorization effect [14]. Ashok [15] et al. compared GloVe with Word2vec (Word to Vector), and experiments proved that using GloVe has better text feature construction capabilities. The research on text feature extraction methods mainly focuses on deep learning, such as Stacked Variational Autoencoder (SVAE) [16], Convolutional Neural Network (CNN) [17-18], Long Short-Term Memory Network (LSTM) [19], etc. Many documents have proved the superiority of deep learning in text feature extraction.

Currently, most text data comes from online social platforms or news portals. When processing this type of text data, it usually needs to perform a series of preprocessing to remove noise information and retain the useful features of the text. The text information input into the neural network is represented by a one-dimensional vector, as shown in Figure 2. Each element in the vector corresponds to the integer identifier of the corresponding word in the text. Word embedding maps semantics into a geometric space by assigning a vector to each word. The distance between any two vectors in the space can be expressed as the semantic relationship between the corresponding words. This model uses the GloVe word embedding method for word embedding. The GloVe word embedding model learns word vectors through the co-occurrence matrix of words in the corpus. It makes the word vector dot product equal to the logarithm of the word co-occurrence probability and uses weighted least squares to represent [20], as shown in formula (1).

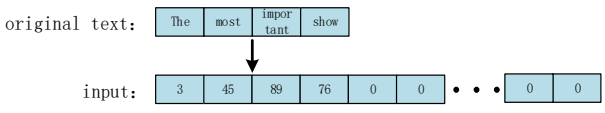

Figure 2. Text vectorization

$$J = \sum_{i,j}^{N} f(X_{i,j})(v_i^T v_j + b_i + b_j - \ln(X_{i,j}))^2. \quad \textbf{(1)}$$

Where X is the co-occurrence matrix, and $X_{i,j}$ represents the number of times that the word i and the word j appear together in a window. v_i and v_j are the word vectors for the word i and j, respectively. b_i and b_j are bias terms. N is the co-occurrence matrix dimension. f is the weight function. The weight function f should meet the following requirements:

(1) When the number of word co-occurrences is 0, the weight is also 0, that is $f(0) = 0$.
(2) The more words co-occur, the weight will not decrease accordingly. $f(x)$ is a non-decreasing function.
(3) When vocabulary co-occurs too frequently, there should not be too much weight.

Pennington [14], the creator of the GloVe algorithm, proposed the following weight function:

$$(x) = \begin{cases} \left(\frac{x}{x_{max}}\right)^{\alpha}, x < x_{max} \\ 1, x \geq x_{max} \end{cases} \quad \textbf{(2)}$$

The weight function works best when $x_{max} = 100$ and $\alpha = 0.75$.

To further obtain the features of text data at different levels of abstraction, the model uses CNN to extract features from the word vector matrix. The word vector matrix M performs convolution operation through the convolution layer to obtain the feature map C. Then it obtains the maximum value $\hat{C}$ in the feature map through the maximum pooling layer. The output feature of the word vector matrix Z_{text} is obtained under the action of multiple convolution kernels, as shown in formulas (3) to (6).

$$C_i = f(wM + b). \quad \textbf{(3)}$$

$$C = [C_1, C_2, \ldots, C_n]. \quad \textbf{(4)}$$

$$\hat{C} = \max\{C\}. \quad \textbf{(5)}$$

$$Z_{text} = [\hat{C}_1, \hat{C}_2, \ldots, \hat{C}_m]. \quad \textbf{(6)}$$

Where n is the number of feature maps, m is the number of convolution kernels.

3.2 Time Series Modeling Based on Fully Connected Network

The deep learning methods commonly used for time series prediction include Recurrent Neural Network (RNN), LSTM network, Fully Connected Network (FC), etc. Among them, LSTM has become the most used deep learning algorithm for time-series prediction due to its ability to obtain medium and long-term dependence on time series data. However, if the time series data is detrended to eliminate the long-term trend of the time series data, the need to obtain long-term dependencies can be largely eliminated. This makes LSTM inferior to FC on some time series forecasting problems.

The fully connected network (FC) is a non-linear model, which can fit non-linear functions well. The input vector of FC adopts the form of lagging observation as the fixed input vector, as shown in formula (7).

$$Z_{ts} = f(wt + b). \quad \textbf{(7)}$$

Where $t = (t_1, t_2, \ldots, t_L)$, L is the lag length.

3.3 Heterogeneous Data Feature Fusion

At present, most researchers use feature splicing and fully connected layers to fuse features in the research of data fusion based on deep learning. Although this type of fusion method is relatively simple, it cannot fully tap the correlation between multi-source heterogeneous data. In this paper, SAE is used as a fusion model to fully mine the relationship between time-series data and text data while retaining the maximum amount of information to obtain the shared feature representation of the data.

Sparse autoencoder is an unsupervised deep learning algorithm, First, the SAE model takes the combined feature x (x contains text data feature Z_{text} and time-series feature Z_{ts}) as input. Then, it encodes the combined feature x as a hidden layer feature representation y. Finally, the hidden layer feature representation y is mapped back to the input space x'. The above steps are called encoding and decoding, expressed as formula (8):

$$\begin{cases} x = concat(Z_{text}, Z_{ts}) \\ y = h(Wx + b) \\ x' = h'(W'y + b') \end{cases} . \quad (8)$$

Where h and h' are the activation function of the encoder and the decoder, W and W' are the weight matrices of the encoder and the decoder, b and b' are the bias of the encoder and the decoder. SAE optimizes the parameters by minimizing the loss function. The loss function of the SAE model is expressed as formula (9):

$$\begin{cases} cost = \frac{1}{n}\sum_K (x_k - x'_k)^2 + \frac{\lambda}{2}\sum_i^L \sum_j^n \sum_i^k \left((w)_{ij}^l\right)^2 + \beta\Psi_{\text{sparty}} \\ \Psi_{\text{sparty}} = \sum_i \rho log\frac{\rho}{\rho_i'} + (1-\rho) log\frac{1-\rho}{1-\rho_i'} \\ \rho_i' = \frac{1}{n}\sum_{j=1}^n y_i(x_j) \end{cases} . \quad (9)$$

Where L is the number of hidden layers. n is the number of data samples. k is the input vector dimension, λ and β are given coefficients, which control the weight coefficient regular term and sparse regular term respectively. $\rho_i^{'}$ is the average activation value of hidden layer neurons and ρ is the sparsity parameter.

4 Model Algorithm Steps

The multi-source heterogeneous data fusion model based on FC-SAE proposed in this paper aims to solve the difficult problem of multi-source heterogeneous data fusion. Through feature extraction of multi-source heterogeneous data, and using the SAE model for fusion, the purpose of making full use of multi-source heterogeneous data is achieved. The FC-SAE model algorithm flow is shown in Figure 3. The specific steps are as follows:

Step 1: Delete HTML tags, delete stop words, lowercase conversion, and other preprocessing of text data.

Step 2: Use the Tokenizer to segment the text, generate a variable-length text sequence, and perform zero-padding to become an equal-length text vector $V_T (V_T = R^{n\times S}$, n is the number of text data samples, S is the maximum text vector length).

Step 3: Use the GloVe word embedding model to embed the text vector V_T to generate a word vector matrix $M (M = R^{S\times n\times i}$, i is the dimension of the word embedding matrix).

Step 4: CNN extracts the features of the word vector matrix M and obtains the deep features Z_{text} of the text data.

Step 5: Preprocess the taxi travel data and weather data, generate event existence data D_E based on the event description data.

Step 6: Use the historical average method to detrend the taxi trip data, as shown in formula (10).

$$\begin{cases} T = TS - \overline{TS} \\ \overline{TS} = \frac{1}{N}\sum_{i=1}^N TS_i. \\ N = 7 \end{cases} \quad (10)$$

Step 7: Use FC neural network to extract features Z_{ts} of taxi travel data.

Step 8: Combine the weather data D_W, the event existence data D_E, and the heterogeneous data features (Z_{text} and Z_{ts}), namely $Z = concat(D_W, D_E, Z_{text}, Z_{ts})$. The combined features Z are fused through the SAE model, and the multi-source heterogeneous data feature y is output.

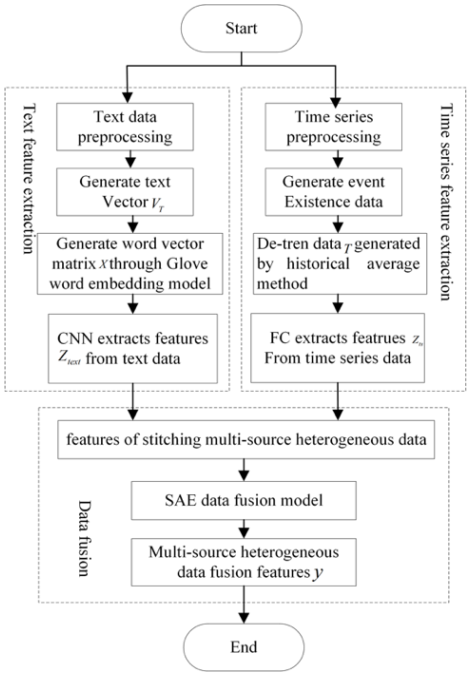

Figure 3. Flow chart of FC-SAE fusion mode

5 Experimental Design and Result Analysis

5.1 Data Set Description and Preprocessing

This paper selects the taxi trip records of the Barclays Center in Brooklyn from the public data set of the New York City Taxi and Limousine Commission as the research object. The weather data is a public data set of the National Oceanic and Atmospheric Administration (NOAA). The weather data set has 13 features, including date, maximum temperature, minimum temperature, wind speed, gust wind speed, visibility, air pressure, precipitation, snowfall, whether it is hail weather, whether it is freezing weather, whether it is foggy, whether it

is thunder and lightning. The text data is obtained from the official website of the Barclays Center, which is a description of the events to be held by the Barclays Center. The text data includes 751 event description text data, each of which includes title, time, and event description. The event existence data is generated from the event text data. The value is 1 when an event occurs, and the corresponding value is 0 on the date when no event occurs. This article selects a total of 637 data in the data set from January 2013 to September 2014 as the training set, selects a total of 91 data from October 2014 to December 2014 as the validation set and the test set selects a total of 564 data from January 2015 to June 2016. An example of the above data set is shown in Table 1.

In the preprocessing of the taxi traveling record data set, the original data is added according to the date to obtain the daily taxi traveling data. The data is detrended to reduce the impact of data offset on calculations. So that the model focuses on the fluctuation of the time series itself and improves the forecasting performance. The preprocessing operation of weather data is to replace the outliers in the data with 0 and normalize the data. The preprocessing of text data is conventional text preprocessing, including HTML tag deletion, lowercase conversion, rooting, and deletion of stop words and prepositions.

Table 1. Sample dataset

(a) Time series data

Date	Hour	Minute	Pickups
2013/1/1	0	0	24
2013/1/1	0	1	42
…	…	…	…

(b) Weather data

Date	Min_te	…	s_ice	ts
2013/1/1	33.1	…	0	0
2013/1/2	21.9	…	0	0
…	…	…	…	…

(c) Text data

Title	Description
Brooklyn Nets vs Houston Rockets	Jeremy James Harden make their only appearance when Houston rocket travel what should high octane arrival revamped rocket have been racing all season net more provides style
New York Islanders vs. Chicago Blackhawks	make dinner reservation today calling emailing email protected click hereto see event menu club restaurant American Express
Brooklyn Nets vs. Milwaukee Bucks	nan
… …	… …

5.2 Experimental Design

The FC-SAE model uses the GloVe word embedding model and CNN neural network for feature extraction of text data. In this model, CNN contains 3 convolutional layers and 3 maximum pooling layers. And it uses the $ReLU$ function as the activation function. A two-layer FC neural network is used to model the time series including 100 neurons and 50 neurons. It uses the $tanh$ function as the activation function. We apply Batch Normalization before every FC layer input. The fusion model uses the SAE model, which includes a 3-layer encoder and a 3-layer decoder. It uses the $tanh$ function as the activation function. The model parameters are optimized by the Adam optimizer.

To verify the superiority of the FC-SAE model compared to other time series forecasting methods in forecasting, this paper compares the FC-SAE model with the commonly used nonlinear time series forecasting model. In addition, the time series modeling module in the FC-SAE model was replaced with LSTM, namely the LSTM-SAE model. In order to evaluate the contribution of different information sources to the fusion features, this paper also conducts an incremental analysis experiment on the FC-SAE model. In order to compare the influence of different fusion strategies on the prediction results, this paper compares the FC fusion strategy and SAE fusion strategy in the case of a complete model.

The experiment uses the platform: Ubuntu 18.04, Intel CPU E5-2620, NVIDIA GTX 1080Ti GPU, and uses TensorFlow-based deep learning library Keras for model construction.

5.3 Model Evaluation

This paper selects Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Goodness of Fit (R^2) to evaluate the proposed model. The calculation formula is shown in the formula (11) to formula (13):

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|y_n - \hat{y}_n|. \tag{11}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}. \tag{12}$$

$$R^2 = 1 - \frac{\sum_{n-1}^{N}(y_n - \hat{y}_n)^2}{\sum_{n-1}^{N}(y_n - \bar{y}_n)^2}. \tag{13}$$

Where N represents the number of samples in the test set, y_n and $\hat{y}_n$ represent the true value and predicted value of the n-th sample, and $\overline{y}$ represents the average value of y_n. MAE and RMSE reflect the degree of deviation between the predicted values and the true values, and their smaller values indicate that the model is more predictive. RMSE is more sensitive to outlier data. R^2 is used to measure the fit of the regression model, and the larger the indicator, the better the fit of the model.

5.4 Model Comparison

5.4.1 Comparison of FC-SAE Model with Other Time Series Forecasting Methods

The prediction results of the FC-SAE model are compared with Support Vector Regression (SVR), Gaussian Process (GP), Long Short-term Memory Network (LSTM), and LSTM-SAE model. SVR uses a linear kernel as the kernel function and sets the penalty factor to 100. GP uses squared exponent and white noise covariance. The LSTM model is a single-layer model, and the number of nodes is 20 and uses just the last time step. In the LSTM-SAE model, the LSTM module is consistent with the LSTM model. The prediction results are shown in Figure 4 and Table 2.

Figure 4 shows that the above models can effectively predict the demand for taxis, but the prediction results are quite different. Through comparison, it is found that the prediction curve of the LSTM neural network and the GP model is quite different from the true value curve, and the effect of fitting the true value curve is poor. Compared with the LSTM neural network and the GP model, the prediction results of the SVR model fit the true value curve better. Compared with the other three models, the LSTM-SAE model and the FC-SAE model have significantly improved the degree of fit between the prediction curve and the true value curve. It can be found that the prediction curve of the FC-SAE model has less fluctuation and can better fit the true value curve compare with the LSTM-SAE model.

Table 2 shows that LSTM is not the best choice in the time series forecasting problem of this paper. The performance of the FC-SAE model cannot be achieved by adjusting the parameters of the LSTM-SAE model. Compared with the SVR algorithm, the GP algorithm, and the LSTM algorithm, the FC-SAE model reduces by 1.3%, 5.3% and 6.6% on MAE, 1.8%, 5.4%, and 4% on RMSE. It increases by 1.8%, 6.6%, and 7.7% on R2 respectively. The results show that the FC-SAE model has obvious advantages. From the experimental results in Table 2, the FC-SAE model and the SVR model are the methods with better performance in this research field, and the time series forecasting method of data fusion has obvious advantages compared with the traditional time series forecasting method.

Table 2. Comparison of FC-SAE model with common time series model and LSTM-SAE model

Methods	MAE	RMSE	R^2 (x100)
SVR	97.2	138.3	61.5
GP	101.2	143.6	58.7
LSTM	102.7	141.5	58.1
LSTM-SAE	98.3	139.7	61.1
FC-SAE	**95.9**	**135.8**	**62.6**

5.4.2 Multi-source Heterogeneous Data Fusion Comparative Experiment

This part of the experiment is carried out by incremental analysis to analyze the influence of different multi-source heterogeneous data on the fusion result. First, only model the time series (TS), and then sequentially add weather data (TS+W), event existence data (TS+W+E), and event text description data (TS+W+E+TE). The results are shown in Figure 5, Table 3, and Table 4.

Figure 5 shows that when time-series data (TS), weather data (W), event existence data (E), and event text data (TE) are added in sequence, the fitting effect of the FC-SAE model's prediction curve and the true value curve is gradually improved. When the event data is added (TS+W+E), the prediction curve fit of the FC-SAE model improves the most. The experimental results show that the error of the time series prediction results will be further reduced. And the prediction accuracy will be further improved when considering the impact of unexpected events, weather, and other factors.

From Table 3 and Table 4, it can find that the introduction of weather data (TS+W) can improve the time-series forecast results. From an empirical analysis, bad weather conditions will change the way people travel. It is consistent with our experimental results. The event existence data (E) that has the greatest impact on the prediction results, after introducing the event existence data, the MAE and RMSE indicators of the FC-SAE algorithm are reduced by 17.1% and 12.6%. Respectively, the MAE and RMSE of the LSTM-SAE algorithm are also reduced by 17% and 15.3%. In addition, this paper uses the GloVe word embedding model and convolutional neural network to introduce event text description into time series forecasting. After combining the event text description data (TE) with other data, the MAE of the FC-SAE model decreased from 99.3 to 95.9, the RMSE decreased from 140.3 to 135.8, and R^2 increased from 0.599 to 0.626. The MAE and RMSE of the LSTM-SAE model are reduced by 6.6% and 2.2%, respectively. Through analysis, it can be known that the introduction of different data has different effects on the prediction results. Through the fusion of correlated heterogeneous data, the prediction errors are reduced effectively.

Table 3. FC-SAE multi-source heterogeneous data fusion comparative experiment

Items	MAE	RMSE	R^2
TS	120.8	166.7	46.7
TS+W	119.8	160.5	47.5
TS+W+E	99.3	140.3	59.9
TS+W+E+TE	**95.9**	**135.8**	62.6

Table 4. LSTM-SAE multi-source heterogeneous data fusion comparative experiment

Items	MAE	RMSE	R^2
TS	128.6	169.8	40.1
TS+W	127	168.5	42.3
TS+W+E	105.3	142.8	59.6
TS+W+E+TE	**98.3**	**139.7**	**61.2**

5.4.3 Fusion Strategy Comparison Experiment

This paper also compares the performance of the SAE fusion strategy and FC fusion strategy to verify the effectiveness of the SAE fusion model. The results are shown in Table 5. Experimental results show that the use of the SAE fusion strategy has a significant improvement in the prediction results compared to the use of the FC fusion strategy. Compared with the FC-FC model, the MAE and RMSE of the FC-SAE model decreased by 1.7% and 2.3%. Compared with the LSTM-FC model, the MAE and RMSE of the LSTM-SAE model decreased by 2.1% and 2.2%. Through analysis, it is found that the advantage of the SAE fusion strategy is that it

can fully explore the correlation between time series data and text data while retaining the largest amount of information. The experimental results verify that the SAE fusion strategy performs better in heterogeneous data fusion than the FC fusion strategy. The SAE fusion strategy can extract effective features and filter redundant information.

Table 5. Comparison of LSTM-SAE and FC-SAE model fusion methods

Methods	MAE	RMSE	R^2
FC-SAE	**95.9**	**135.8**	**62.6**
FC-FC	97.6	139.1	62.2
LSTM-SAE	98.3	139.7	61.1
LSTM-FC	100.4	142.9	59.5

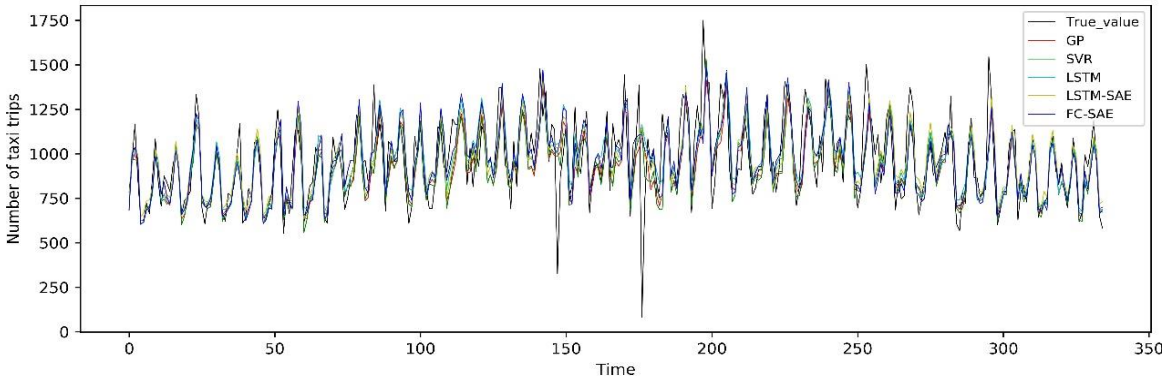

Figure 4. Comparison of prediction results of five models

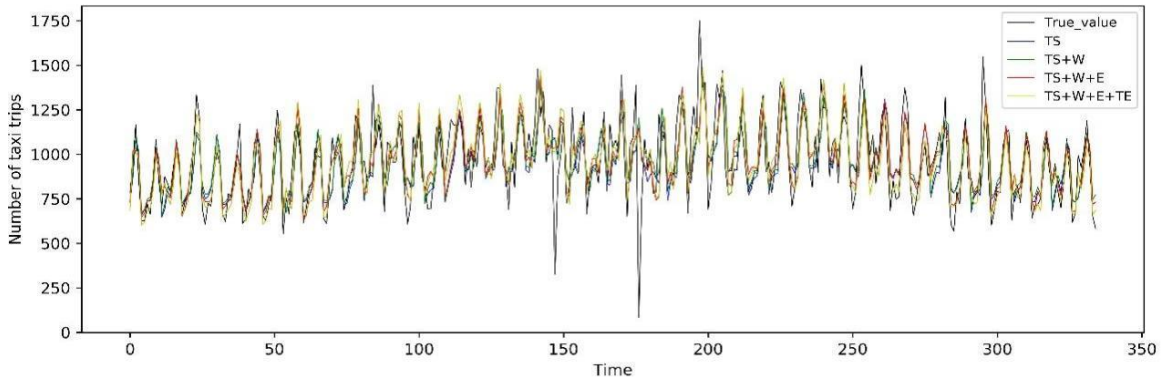

Figure 5. FC-SAE model heterogeneous data fusion comparison

6 Conclusion

Data fusion is one of the important technologies for big data information mining. This paper uses multi-source heterogeneous data as the data source and establishes a multi-source heterogeneous data fusion model based on FC-SAE. The model uses Glove for word vectorization and uses CNN to extract features. It uses multi-layer FC neural network to model nonlinear time series. Then it uses SAE to fuse the event data, weather data, and time series data. This paper proposes a new fusion approach. Although we apply it to the transportation domain to solve the taxi demand problem. The FC-SAE model effectively utilizes the correlation between multi-source heterogeneous data. It fully excavates the fusion feature representation of heterogeneous data and uses the fusion feature to perform time series prediction to obtain more accurate prediction results. Compared with other popular time series forecasting methods, it is proved that the FC-SAE model can reduce the forecasting error well. Multi-source heterogeneous data fusion experiments have proved that multi-source heterogeneous data can effectively reduce the error of time series prediction results and different data have different effects on the prediction results. In addition, the fusion strategy experiment shows the influence of different fusion strategies on the prediction results. The results show that the SAE fusion strategy has better fusion performance than the FC fusion strategy.

This paper only considers the issue of using deep learning to fuse unstructured text data, structured numerical data and time-series data. In the future, we will include other forms of data, such as image, video, and audio, to further research the problem of multi-source heterogeneous data fusion.

7 Funding

This work was supported by National Key R&D Program of China (No. 2019YFB1707303).

References

[1] N. van Oort, T. Brands, E. de Romph, Short-Term Prediction of Ridership on Public Transport with Smart

Card Data, *Transportation Research Record*, Vol. 2535, No. 1, pp. 105-111, January, 2015.
[2] Z. Alzamil, D. Appelbaum, R. Nehmer, An Ontological Artifact for Classifying Social Media: Text Mining Analysis for Financial Data, *International Journal of Accounting Information Systems*, Vol. 38, Article No. 100469, September, 2020.
[3] J. Schmidhuber, Deep Learning in Neural Networks: An Overview, *Neural Networks*, Vol. 61, pp. 85-117, January, 2015.
[4] J. Liu, T. Li, P. Xie, S. Du, F. Teng, X. Yang, Urban Big Data Fusion based on Deep learning: An Overview, *Information Fusion*, Vol. 53, pp. 123-133, January, 2020.
[5] T. Meng, X. Jing, Z. Yan, W. Pedrycz, A Survey on Machine Learning for Data Fusion, *Information Fusion*, Vol. 57, pp. 115-129, May, 2020.
[6] L. Zhang, H. Gao, J. Wen, S. Li, Q. Liu, A Deep Learning-Based Recognition Method for Degradation Monitoring of Ball Screw with Multi-Sensor Data Fusion, *Microelectronics Reliability*, Vol. 75, pp. 215-222, August, 2017.
[7] H. Chen, N. Hu, Z. Cheng, L. Zhang, Y. Zhang, A Deep Convolutional Neural Network based Fusion Method of Two-Direction Vibration Signal Data for Health State Identification of Planetary Gearboxes, *Measurement*, Vol. 146, pp. 268-278, November, 2019.
[8] J. Wu, K. Hu, Y. Cheng, H. Zhu, X. Shao, Y. Wang, Data-Driven Remaining Useful Life Prediction Via Multiple Sensor Signals and Deep Long Short-Term Memory Neural Network, *ISA Transactions*, Vol. 97, pp. 241-250, February, 2020.
[9] E. J. Ruiz, V. Hristidis, C. Castillo, A. Gionis, A. Jaimes, Correlating Financial Time Series with Micro-Blogging Activity, *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, Seattle, Washington, USA, 2012, pp. 513-522.
[10] F. C. Pereira, F. Rodrigues, M. Ben-Akiva, Using Data from the Web to Predict Public Transport Arrivals under Special Events Scenarios, *Journal of Intelligent Transportation Systems*, Vol. 19, No. 3, pp. 273-288, 2015.
[11] L. Li, F. Zhu, H. Sun, Y. Hu, Y. Yang, D. Jin, Multi-Source Information Fusion and Deep-Learning-Based Characteristics Measurement for Exploring the Effects of Peer Engagement on Stock Price Synchronicity, *Information Fusion*, Vol. 69, pp. 1-21, May, 2021.
[12] X. Ye, H. Dai, L. A. Dong, X. Wang, Multi-View Ensemble Learning Method for Microblog Sentiment Classification, *Expert Systems with Applications*, Vol. 166, Article No. 113987, March, 2021.
[13] M. Zulqarnain, R. Ghazali, M. G. Ghouse, M. F. Mushtaq, Efficient Processing of GRU based on Word Embedding for Text Classification, *JOIV: International Journal on Informatics Visualization*, Vol. 3, No. 4, pp. 377-383, December, 2019.
[14] J. Pennington, R. Socher, C. D. Manning, Glove: Global Vectors for Word Representation, *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, Doha, Qatar, 2014, pp. 1532-1543
[15] A. Ashok, R. Elmasri, G. Natarajan, Comparing Different Word Embeddings for Multiword Expression Identification, *International Conference on Applications of Natural Language to Information Systems*, Salford, UK, 2019, pp. 295-302
[16] L. Che, X. Yang, L. Wang, Text Feature Extraction based on Stacked Variational Autoencoder, *Microprocessors and Microsystems*, Vol. 76, pp. 103063, July, 2020.
[17] A. K. Sharma, S. Chaurasia, D. K. Srivastava, Sentimental Short Sentences Classification by using CNN Deep Learning Model with Fine Tuned Word2Vec, *Procedia Computer Science*, Vol. 167, pp. 1139-1147, 2020.
[18] S. Wang, Y. Rao, X. B. Fan, J. N. Qi, Joint Event Extraction Model based on Multi-feature Fusion, *Procedia Computer Science*, Vol. 174, pp. 115-122, 2020.
[19] M. Sundermeyer, R. Schlüter, H. Ney, LSTM Neural Networks for Language Modeling, *Thirteenth Annual Conference of the International Speech Communication Association*, Portland, Oregon, USA, 2012, pp. 194-197.
[20] A. Onan, Mining Opinions from Instructor Evaluation Reviews: A Deep Learning Approach, *Computer Applications in Engineering Education*, Vol. 28, No. 1, pp. 117-138, January, 2020.

Biographies

Hong Zhang received the Ph.D. degree in system engineering from Lanzhou University of technology in 2018. Her main research fields are machine learning, big data analysis, and intelligent transportation.

Kun Jiang is a graduate student in the school of computer and communication engineering, Lanzhou University of technology. His research direction is data fusion.

Chuanqi Cheng received the M.S. degree from Lanzhou University of technology in 2020. His research direction is data fusion.

Jie Cao is currently a member of the Standing Committee of the Party committee and vice president of Lanzhou University of technology. Her main research interests are machine learning, fault diagnosis, and the theory and application of intelligent transportation systems.

Wenyue Zhang is currently studying at the School of Computer and Communication, Lanzhou University of Technology. Her research direction is data fusion.