Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230165

Research on Forecasting Sales of Products Based on Spatiotemporal Graph Neural Network

Shao-yan ZUO[a], Hai-fu HUANG[a], Lin GAO[b,1], Hai-ying LI[c], Hai-bo WANG[c], Chen ZHANG[c], Zhang-sheng WU[c]

[a]*China Tobacco Guangxi Industrial Co., Ltd., Guangxi 530001, China*

[b]*Shandong Qingdao Tobacco Co. LTD, Shandong 256400, China*

[c]*Zhongke Zhidao Technology Co., LTD., Beijing 100190, China*

Abstract. Intelligent marketing and recommendation are a core business of commercial companies, and accurate prediction of sales is the premise and foundation for greater efficiency of smart marketing and recommendation. In order to predict product sales, deep neural network (DNN), convolutional neural network (CNN), time series analysis and other methods have been put forward, but most of which only focus on the temporal or spatial characteristics of data. According to modeling and analyzing sales of products, they are closely related to the spatial location and time of the corresponding merchants. The goal is to predict the sales of products accurately at a given time and place, we advance a hybrid model of CNN-LSTM to forecast sales. Firstly, a large-scale knowledge graph system based on merchants is constructed, which describes the sales data and the relevant interaction scenarios of the corresponding business, merchants and users through the data model of a graph, and add the spatial and data characteristics of the business data on the graph model to describe the temporal and spatial characteristics of the merchants. Based on the constructed business knowledge graph, graph convolutional neural network (GCN) is used to aggregate information and obtain spatial features. Correspondingly, long short-term memory (LSTM) is used to extract time features. Researchers combine the two characteristics to make the sales forecast. In this study, neural network and GCN-LSTM algorithm are respectively used to carry out experiments on two kinds of product regulations. The result shows that the sales predicted by hybrid model of GCN-LSTM is almost as equal as the actual sales. The average accuracy of the proposed model is 89%.

Keywords. Graph neural network, Sales prediction, Long-short term memory, Product, Big data

1. Introduction

Marketing and recommendation are the core of commercial companies as well as the premise prediction for accurate sales. In the past, commodity marketing and recommendation were characterized by a lot of man-power, material and financial resources due to a huge amount of business. At the same time, the repetitive and heavy

[1] Corresponding Author: Lin Gao, Shandong Qingdao Tobacco Co. LTD, Shandong 256400, China. Email: watermelongl@qq.com, 270377937@qq.com

business have encountered some difficulties in the market. Therefore, we propose an intelligent method to predict and recommend some products to meet the needs of merchants so as to predict sales accurately. Many researchers [1-6] devote themselves to it, for instance, LSTM network model was built which was convenient for data input with great accuracy in time series prediction. A cost-aversion biased combination prediction model was proposed on the basis of random forest, GBDT and XG-Boost algorithms, which could accurately predict sales. The time series model based on multi-population genetic algorithm proposed that showed a high prediction accuracy for commodities. Deep learning prediction models combined with pictures and other structured data are proposed to advance a more accurate sales prediction method. However, there are few systematic studies on the combination of Spatio-temporal characteristics of product marketing. Graph convolution network (GCN) and long short-term memory (LSTM) spatiotemporal network were used for rolling frequency prediction [7]. We construct a merchant knowledge map to represent the relationship among stores. And GCN-LSTM hybrid model is constructed to extract spatial features and temporal features, which will be combined with sales prediction. We raise this method that can obtain more ideal marketing forecast.

2. Overview of research

Forecasting commodity marketing is one of the most important businesses among commercial companies. Traditional forecasting methods inevitably have heavy data workloads and manual limitations in fetching, memory, calculation, etc. It is difficult to achieve a large-scale accurate calculation, which will directly affect the economic benefits of commercial companies and merchants. Therefore, artificial intelligence methods for accurate predictive marketing have been the key research direction of various companies.

Deep neural network has a powerful feature extraction [8], graph convolutional neural network (GCN) has shown a strong application performance as a deep learning representation algorithm, which not only express complex semantic relations, but also capture global graph information [9]. LSTM is widely used in the field of artificial intelligence, multiple LSTM units forming an LSTM network can be used to learn the characteristics of the temporal dimension of the input spatiotemporal data.

In order to predict sales of commodities accurately, different data types of commodities are used as input. In this paper, a new framework based on graph convolutional neural network (GCN) and long short-term memory (LSTM) neural network is proposed for the prediction of time and space sales. Specifically, the main work of this study is organized as follows:

(1) Perform feature fusion and process data, complement 0 for numerical data, and perform log1p function to obtain smooth data relatively, and class data is supplemented with null, encoded by Label-Encoder, and converted to obtain numerical features. In neural network A. kaiming-normal is used for weight initialization and other operations. After the fusion of features, the data is input into neural network B and passed through ReLU respectively. Batch normal layer normalization and dropout operation are processed to obtain the difference between the measured target value and the predicted value.

(2) The prediction method of graph neural network is proposed. Construct many store knowledge graphs in the business district, extract node and the edge of the attribute

to predict relationships between a node and an edge, after making a knowledge fusion, Graph convolution neural network algorithm is proposed. We establish knowledge graphs to input data, neighbors of each node are used by convolution operation, convolution results update the node, realize information aggregation, the nonlinear activation function is used for two-layer convolution to achieve the expected depth.

(3) A GCN-LSTM hybrid model is constructed for the prediction of product sales by graph convolutional neural network. LSTM is used to solve the problem that long-term dependent feature relations cannot be learned. The proposed model fuses GCN related spatial data and LSTM time data to create a new data, which makes full use of the location and the structure of shops, LSTM capture the characteristics of the dynamic sales change from several aspects of time correlation to achieve the ultimate for goods sales forecast, and improve the accuracy of commodity sales prediction.

(4) In order to express the accuracy about intelligent prediction of product sales, two kinds of products with different product regulations are carried out relevant visual experimental verification of sales prediction in different time periods. In contrast to actual sales, it is shown that prediction accuracy of GCN-LSTM model is 89% on the average.

3. Algorithm of Sales forecasting based on shallow neural network

With the progress of development of information technology, more and more methods can be used to deal with the prediction such as BP neural network, time series analysis prediction, prediction based on support vector machine, etc. Although these methods have good effects in some aspects, they still have many shortcomings. However, neural networks in deep learning can effectively solve the limitations of traditional methods, deal with the prediction tasks with high complexity, and the prediction results are closer to the real value. Therefore, this study proposes a product delivery algorithm based on shallow neural network, which is used to predict product sales and achieve a more accurate marketing forecast of products. The flow chart of product delivery algorithm based on shallow neural network proposed in this study is shown in Figure 1.

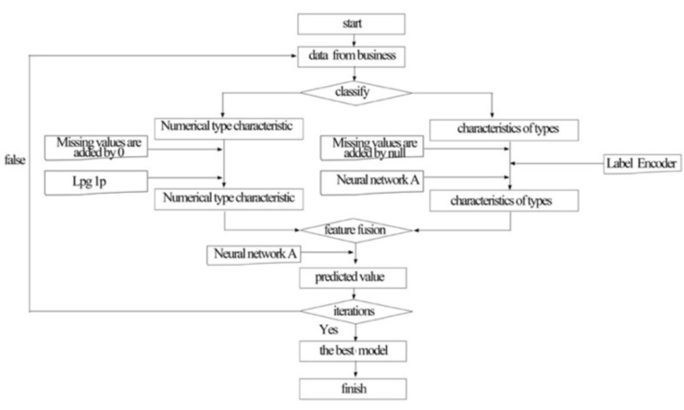

Figure 1. Flow chart of product delivery algorithm based on shallow neural network

Label Encoder transforms the categorical data into numerical data. If there is a label of a mobile phone brand, under which there are "Apple", "Huawei" and "Xiaomi", Label Encoder will be called to map "Apple", "Huawei" and "Xiaomi" respectively from 0, and the result is 0, 1, 2. The structure of neural network A is shown in Figure 2.

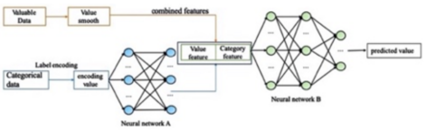

Figure 2. Neural network architecture diagram.

Figure 3. The knowledge graph of Store

Neural network A is a two-layer neural network, and Kaiming- normal which is a particularly robust initialization method that adapts to the nonlinear activation function is used for weight initialization. After the output of the first layer neural network, batch-normal function is used to normalize the output to the standard distribution shape, and ReLU is used to activate the output. After the output of the second layer neural network is inserted, the output of the second layer neural network is regarded as a set of category features generated by the neural network after learning. The smoothed numerical feature will be fused with the category feature. Since both the numerical feature and the neural network feature are vectors, this algorithm spliced the two feature vectors, as shown in Formula (1), where NFi represents the numerical feature, EFi represents the category, and CFi represents the fusion feature. NFi can be interpreted as $NFi^T = (NF_1, NF_2 \dots\dots, NF_n)$, $EFi^T = (EF_1, EF_2 \dots\dots, EF_m)$,whereas CFi is m plus n vector.

$$CF_i = [NF_i, EF_i] \tag{1}$$

After the features are fused, they are fed into neural network B. The neural network is a three-layer structure, and kaiming-normal is also used for weight initialization. In the output of the first layer, the ReLU is used to activate the dropout operation, and the output is sent to the batch-normal layer for normalization. Then, the output of the second layer is sent to the God network, and the output of the second layer is also performed a dropout operation. Finally, the output of the fully connected layer is connected and the output of the fully connected layer is used as the predicted value. The entire neural network is trained using MSE Loss, which measures the difference between the target value and the predicted value, as shown in Equation (2), where x is the predicted value and y is the true value.

$$loss(x_i, y_i) = (x_i - y_i)^2 \tag{2}$$

The neural network optimizer uses Adam optimizer with a learing rate of 0.0001, and 64 groups of data are input for algorithm training in each batch. We have tried another optimizer to test the results such as SGDM, SGD, Adagrad, etc. However, we find the Adam shows the best performance in the real application. For your references, we have listed the contrast in table 1.

Table 1. Performance analysis of different optimizers in deep learning model of commodity sales prediction

Optimizer	SGD	Adam	Adagrad	SGDM
Time/s	12.212	12.158	14.213	13.272
Loss	0.4171	0.3511	0.5355	0.4388
accuracy	0.8451	0.8964	0.8815	0.8317

4. Convolutional neural network algorithm

4.1 Figure

A graph is a data structure that describes the relationship between the one and the other one. It is usually composed of nodes and edges, where nodes represent entities, and edges between nodes indicate that there is relationship between two entities. A graph can be also expressed as $G = \{V, \mathrm{E}\}$, where V is the set of all nodes in the graph, E represents an edge set that describes the relationship between the nodes and the adjacent node. As Formula (3) shown, the graph has many nodes, $A = (a_{ij})_{n\times n}$,which represents the adjacency matrix of the graph.

$$a_{ij} = \begin{cases} 1\ , v_i\ and\ v_j\ \ are\ connected \\ 0, v_i\ and\ v_j\ aren'tconnected \end{cases} \tag{3}$$

If graph is an undirected graph, $a_{ij=}a_{ji}$, graph represent the degree matrix, and denotes the number of connected nodes. Assuming that each node in the graph has a feature, the features of all nodes in the graph are combined into a matrix.

4.2 Business area information atlas construction

The knowledge graph includes knowledge modeling, knowledge extraction, knowledge fusion, knowledge storage, knowledge reasoning, and knowledge application [10]. The "entity-relational-entity" is represented by SPO triplet, and the feature information comes from the attributes of nodes and edges. This form has a strong interpretability and can directly reflect the structural and semantic information of the atlas. Based on the store information, the spatial size data collected in this study are shown in Table 2, the knowledge graph constructed is shown in Figure 3, and the overall framework is shown in Figure 4.

Table 2. The capacity of Knowledge graph

Number of the store	Number of products	Node	Edge
4	2	10	9

In the process of building store knowledge graph, the data used are multi-source and heterogeneous. Structured data comes from the integration of research teams. Semi-structured data comes from relevant data of commodity delivery platform, which needs to be normalized by attributes. Unstructured data comes from fragmented text content, and knowledge processing needs to extract information from retail label data in the business area. After the data source is obtained, the knowledge is modeled. It is defined as Conceptual Domain - Entity Domain - Event Domain. The conceptual domain is the abstraction of the concrete entity, and the edge between the two entities is represented by the distance between the stores, the product regulation and the sales volume respectively. Entity domains are business-relevant instances, such as different store names, product specifications, and sales volumes in the graph. The event domain refers to the purchasing behaviors of customers. The events behind these behaviors are used as the precipitation of structured knowledge to enhance some static knowledge in the entity domain.

The third part is knowledge processing. For semi-structured or unstructured business data, the acquisition of knowledge will involve classification or extraction, such as how to define the relationship between each store entity. After knowledge acquisition, the structured data of knowledge is mapped to the architecture in knowledge modeling. Then do the entity chain refers to the normalization work, one is the entity link, two is the entity normalization or attribute normalization. After knowledge processing, triplet data is fused to make knowledge inference on complex semantic relations, and attributes and relations are predicted: two stores less than 3 kilometers can correspond; Shops belong to the area of the business district, and then predict the rules and sales of a product, the dragon (hard Ling-yun) belongs to the sales rules in the store, and its sales volume is 443 boxes.

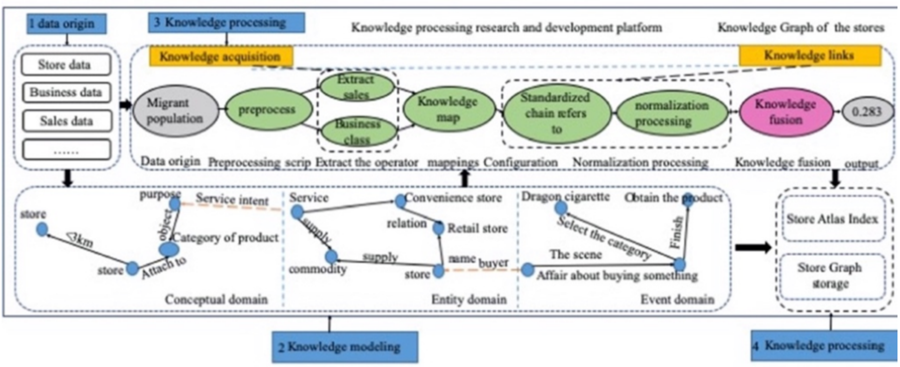

Figure 4. Overall framework of the knowledge graph

4.3 Overview of neural network

Convolutional neural network extracts data features through convolution. However, classical convolutional neural networks can only process Euclidean spatial data such as images and texts among others which are of translation invariance. In order to process non-Euclidean spatial data such as graph data, F. Scarselli have proposed the graph Neural network (GNN) [11] method. GNN can transform graph- structure data and then input them into various neural networks for training. Graph neural network can be classified [12] into Graph Convolution Network, Graph Attention Network, and Scalable Graph Network.

4.3.1 Figure Convolutional Neural network

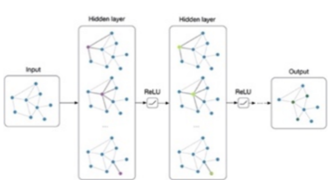

Figure 5. Training process diagram of GCN

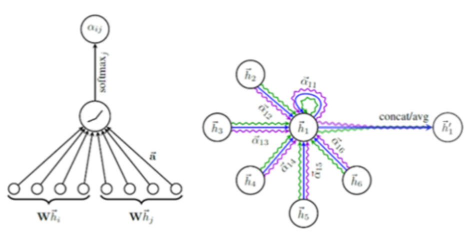

Figure 6. The learning process of attention network

Bruna [13] et al. proposed a method to combine graph and convolutional neural networks i.e. graph convolutional neural network (GCN) in 2013which aims to accomplish node prediction, node classification and link prediction making use of features information extracted from graph-structure data. The training process of GCN is shown in Figure 5 below. GCN takes the graph data as inputs, and performs a convolution on each node's neighbors, and updates the node with the result of convolution to realize the aggregation of information among nodes. The result is then convolved as the second layer input and the above operation is repeated until the number of layers reaches the expected depth. Finally, GCN can also convert node states into task-related labels, etc. as outputs.

There are two ways to extract graph features, one is though spectral domain, the other is though vertex domain. The spectral domain uses the eigenvalues and eigenvectors of the Laplacian matrix of the graph to study its properties, the vertex domain finds out the neighbor nodes adjacent to each vertex, defines the connection relationship between the nodes, and then aggregates the information of the neighbor nodes.

(1) Spectral-based Graph Convolutional Neural Networks

The Laplacian matrix of the graph is written as $L = D - A,$ where D refers to the degree matrix of nodes and A refers to the adjacency matrix of nodes. GCN is based on Laplace's spectral decomposition, GCN decomposes a matrix into the product of eigenvalues and eigenvector matrices, In this equation, U is the orthogonal matrix composed by the unit eigenvectors. As shown in Equation (4).

$$L = U\Lambda U^{-1} = U\Lambda U^{T} \tag{4}$$

The first graph convolution model proposed by Bruna is as follows,but the $y_{output=\sigma(Ug_{\theta}(\Lambda)U^{T}x)}$original graph convolution neural network has a large amount of computation, but it is of poor performance due to massive calculation and lack of spatial locality. In order to increase the spatial locality of the first-generation graph convolution model, De fferard et al. [14] proposed the second-generation graph convolution neural network in 2016 as in Formula (5).

$$y_{output=\sigma\sum_{j=0}^{K} a_j L^j} \tag{5}$$

It not only improved kernel convolution, but also further simplify the calculation. After several improvements since GCN was proposed, T.N.Kipf and M.Welling et al. [15] put forward the third-generation graph convolutional neural network in 2017,which deepened the depth of the network and defined the propagation mode between layers as shown in Formula (6) .

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{\frac{-1}{2}}\tilde{A}H^{l}\widetilde{D}^{\frac{-1}{2}}W^{l}\right) \tag{6}$$

In the equation, $\tilde{A} = A + I, \widetilde{D}$is the degree matrix of$\tilde{A}$,H Is the characteristic of each layer, for the input layer, W is the parameter matrix, σ is the nonlinear activation function.

(2) Vertex-based Graph Convolutional Neural Networks

Graph convolution based on vertex domain can be likened to the convolution operation in Euclidean space. Starting from vertex domain, each central node and its neighbors are aggregated by defining an aggregation function. It can be seen that the problem of vertex domain graph convolution lies in the different number of neighbor nodes of each node in the graph structure. In order to solve this problem, it is necessary to define a convolution kernel that can handle neighbor nodes of arbitrary length, so that the convolution kernel can perform convolution adaptively according to the number of neighbor nodes of different nodes, so as to further extract the features of nodes in the graph. Sum the hidden states of all neighbor nodes, update the current node hidden states, and realize parameter less convolution: the feature representation of the layer, representing the set of neighbor nodes of a node. Graph convolution based on vertex domain can deal with large scale graph structures and is widely used.

4.3.2 Figure Attention network

The method of GCN to obtain the features of the graph space is very dependent on the graph structure, and the weight of different nodes in the neighborhood is the same. In 2018, Yoshua Bengio et al. proposed a Graph Attention network (GAT) model combining attention mechanism and graph convolutional neural network. Graph attention network has two main advantages. One is that different weights can be assigned to each node, and the second is that after introducing attention mechanism, node information is only related to its neighbor nodes, without obtaining the information of the whole graph. Assume that the graph has n nodes and take the set of node feature vectors as input, each node has F features. After being processed, the dimension of the feature vector may change. For both nodes, a linear transformation is used to transform the dimensional features into dimensional features. For a node, calculating the similarity coefficient as $e_{ij} = a\left(W\vec{h}_i, W\vec{h}_j\right)$ and each neighbor node is a shared attention mechanism, which can map the concatenated vector to the real number. Finally, after it is normalized by Soft-max, the attention coefficient can be obtained as shown in Formula (7).

$$a_{ij} = \frac{exp\left(LeakyReLU\left(e_{ij}\right)\right)}{\sum_{k \in N_j} exp\left(LeakyReLU\left(e_{ij}\right)\right)} \tag{7}$$

In the second step, weighted summation of features is carried out, and the feature vector of fusion domain information of each node is output after nonlinear activation function. As shown in Figure 6, the graph attention network learns the attention weight between two nodes, and performs weighted average on the feature representation of nodes based on the attention weight, so that the feature representation of node 1 can be obtained.

4.3.3 Extensible graph network

The convolution done in GCN incorporates the information from the whole Graph and the efficiency of the GCN can be low if the graph structure is large and the number of nodes is high, which leads to the emergence of the scalable graph network Graph-SAGE [16] with the learning process shown in Figure 7 below. The network is sampled by randomly picking subgraphs and updating the nodes through them, in this way the resulting subgraph structure itself is changing, as a result the model will learn a sampling and aggregation of parameters, avoiding the need to update the node features of the whole graph together during training, thus increasing the scalability.

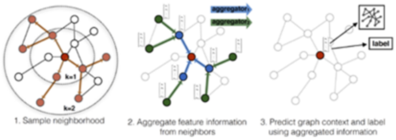

Figure 7. Graph-SAGE learning process

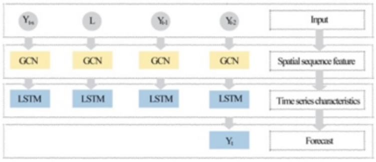

Figure 8. The training flow chart of sales forecasting based on GCN-LSTM

Graph-SAGE first partitions each node to get subgraphs, and then randomly samples some neighbor nodes as feature points for aggregation. After picking out the subgraphs, feature fusion is done, which is able to get the features of the central node in the same way as used by GCN, and finally the nodes are made for classification tasks, etc.

5. GCN-LSTM hybrid model

The sales of goods have time characteristics, and they are related to other spatial characteristics. Graph convolution neural network can be used to process graph data and extract spatial features. LSTM can be used to process time series problems and extract time features of data. Therefore, many scholars have proposed to use the combination of GCN and LSTM to predict traffic flow or air quality, etc. Zhao Ling[17] proposed T-GCN model to predict traffic flow in combination with GCN and GRU, and Qi Bolin [18] used GCN-LSTM model to monitor air quality of small and micro stations. Therefore, we can try to make use of graph convolution neural network model based on the combination of graph convolution neural network and LSTM for sales on the basis of data with both temporal and spatial characteristics.

5.1 Model construction

For commodity sales, if the store is adjacent to the other store, where the sales of the two retailers interact on each other. Therefore, each merchant is considered as a node, connections do exist between merchants if the distance is less than 3km.We try to build a merchant graph, which represents the number of nodes, an edge, and an adjacency matrix, and represents the connectivity between merchants. None connectivity between merchants exist, the distance between the node is zero, the condition is that the distance is 1 from the nearest retailer. Some information related to the sales of goods from the merchant is regarded as the attribute characteristics, forming a feature matrix to represent the number of node characteristics. Node characteristics include the basic attributes of the business circle where the store is located, crowd characteristics, consumption capacity and commodity sales mentioned above. Node attribute characteristics represent time.

Therefore, the GCN-LSTM model learns the following mapping through the graphic structure of the store and the corresponding feature matrix such as predicting future sales. The model is divided into two steps such as GCN and LSTM. The training flow chart is shown in Figure 8 below. The historical data of merchants are as input, GCN aggregation information is used to obtain spatial characteristics, and then input the time series data integrated with spatial characteristics into the LSTM model to capture the time characteristics, we aim at obtaining the final prediction results.

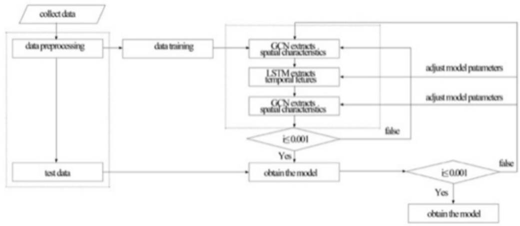

Figure 9. Flow chart about forecasting sales of GCN-LSTM.

The process of GCN-LSTM model about forecasting commodity sales is shown in Figure 9. After preprocessing the input data, the data is divided into training and testing set, If the model does not do well in the test data, we will adjust the parameter of the test model.

5.2 Spatial feature extraction

The key to predict sales is to obtain a spatial structure among businesses. Traditional convolutional neural networks can extract features from spatial data, and deal with the graph structure of locations among businesses. Therefore, graph neural networks are used to process graph data and extract spatial features. GCN can collect the relationship between merchants and their surrounding connected merchants, aggregate the information among merchants, and obtain the spatial correlation of features. In the GCN process, the characteristic matrix containing the merchant information and the adjacency matrix reflecting the location relationship are used as inputs to enter the convolution layer for calculation and aggregation of information. The result of each aggregation will enter the convolution layer again as a new input until the preset number of convolution layers is reached. The final output result is the GCN output that combines the surrounding information and its own information.

5.3 Extraction of time characteristics

Another key to predict the sales of goods is to obtain the time correlation among data. Traditional RNN is limited to long-term prediction and is only suitable for learning short-term memory. As a variant of GNN, LSTM [19] can solve the problems in GNN by using the gating mechanism to remember much long-term information. Therefore, LSTM model is used to obtain time characteristics from sales data, which represents the hidden state of the time, the sales information of the time, the cell state of the time, and the output state of the time.

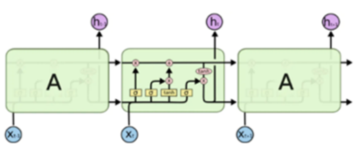

Figure 10. Structure of LSTM [20]

Algorithm 1 The Process of Graph Convolution

Input:
The graph $G(V, E)$;feature matrix X_{t-1};adjacency matrix A;
degree matrix $\hat{D}$;layer number L
$\hat{A} = A + I$
$\hat{D} = \sum_{j=1}^{n} A_{ij}$
for $i = 1$ to n **do**
  for $l = 1$ to L **do**
    $H^0_{t-i} = X_{t-i}$
    $H^{l+1}_{t-i} = \sigma(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^l_{t-i} W^l)$
  end for
  $\hat{X}_{t-i} = H^L_{t-i}$
end for
$X_t = GRU([X_{t-n}, X_{t-n+1}, ..., X_{t-1}])$
Output:
X_t

Figure 11. Pseudo code of graph convolution.

LSTM decides which information is discarded from the cell state through the forgetting gate, what is put into the cell state through the input gate, after updating the cell state, output the final value through the output gate. We input the hidden state of the moment and information about the current sales to predict the sales of the next moment. The structure of LSTM is shown in Figure 10.

5.4 GCN-LSTM

In order to predict the sales, GCN-LSTM model is used to simultaneously extract the spatial and temporal characteristics of the data. The specific calculation process is as follows:

$$Ft=\sigma\left(w_f * \left[h_{t-1,}x_t\right] + b_f\right) \tag{8}$$

$$i_t = \sigma(w_i * \left[h_{t-1,}x_t\right] + b_i \tag{9}$$

$$z_t = tanh(w_c * \left[h_{t-1,}x_t\right] + b_c \tag{10}$$

$$c_t=f_t{}^*c_{t-1}+i_t * z_t \tag{11}$$

$$o_t = \sigma(w_o * \left[h_{t-1,}x_t\right] + b_o \tag{12}$$

$$h_t = o_t * tanh\ (c_t) \tag{13}$$

Wherein, the process of graph convolution, the weight and deviation are in the training process. The pseudo code for the graph convolution process in this study is shown in Figure 11 below. The graph convolution process in GCN-LSTM model can make good use of the location structure among adjacent businesses to obtain the spatial characteristics in the data. The LSTM process can capture the characteristics of the dynamic change of sales with time, obtain the time correlation, and finally achieve the prediction of commodity sales.

6. Experimental analysis and conclusion

The work content of this study can automatically extract important features from data to make up for the heavy workload of manual feature search. Potential features among data that cannot be extracted manually, so as to improve work efficiency and realize intelligent and accurate product marketing. However, there are some shortcomings in this research work, such as incomplete data collection and many comprehensive factors which affect product delivery. In the future, we will also build relevant models according to local conditions to create more profound and effective application value in the field of intelligent product marketing.

6.1 Data description and preprocessing

The data of business circle and sales from different brands may be empty when input. In order to facilitate the subsequent learning of neural network, it is necessary to process them.

Table 3. The label of retailer

Basic attribute of business district	The number of residential areas/The number of shopping centers/The number of office buildings
Features of the crowd	Resident population/The floating population/Working population/The population of permanent residents/Gender/ age /degree /marriage /status of economy
Consumption level of business scope	In the gear/Low bit/The average gear/Level of consumption/ The dining level
Market status of the business area	Average price of from different brands in recent three months
Indicators about business circle consumption	Sales of different brands in recent March
Consumption preferences in the business circle	Weekly sales of different brands of products

The numerical data should be supplemented with 0, and the categorical data should be supplemented with null. Considering the data set is numeric data type and categories, numeric data under different business circle properties could be orders of magnitude more than the gap, could lead to a neural network to ignore small orders of magnitude, the characteristics of the categories of data due to not numeric, not directly input neural network training, so the need for numeric data type and category data transformation. For numerical data, log1p function operation is performed uniformly, and a relatively smooth and Gaussian distribution data is obtained. Label-Encoder is used to encode the categorical data to obtain the numerical features as shown in table 3.

6.2 Evaluation Index

In order to evaluate the prediction performance of the two algorithms for sales, absolute mean error (MAE) and root mean square error (RMSE) are used to evaluate the deviation among actual sales and predict sales, and accuracy is used to evaluate the effect of the model, as shown in Equation (14-16). The better the prediction is, the closer the prediction is to the real value, and the smaller the deviation is, the smaller the MAE and RMSE will be.

$$\mathrm{MAE}=\frac{1}{n}\sum_{t=1}^{n}|y-\check{y}_t| \tag{14}$$

$$\mathrm{RMSE}=\sqrt[2]{\frac{1}{n}\sum_{t=1}^{n}(y-\check{y}_t)^2} \tag{15}$$

$$\mathrm{Accuracy}=1-\frac{\|y-\check{y}_t\|_F}{\|y\|_F} \tag{16}$$

6.3 The experimental results

In the experiment, shallow neural network and GCN-LSTM models are respectively used to predict the cigarette sales of 10 retail households. The comparison of prediction results is shown in figure 12.

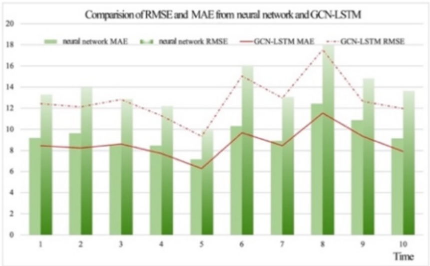

Figure 12. Comparison of RMSE and MAE from neural network and GCN-LSTM

As can be seen from the results in the table, the MAE and RMSE evaluation indexes of the GCN-LSTM model are both smaller than those of the shallow neural network model, indicating that the deviation between the predicted value and the real value of the sales volume of the model is smaller and the prediction effect is better.

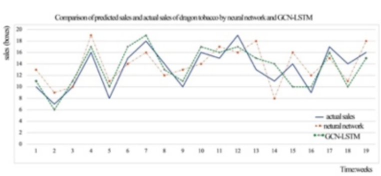

Figure 13. comparison of two methods in 20 weeks

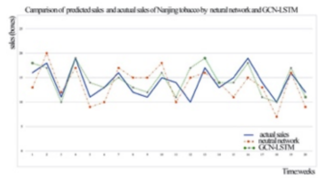

Figure 14. comparison of two methods in 20 weeks

The experiment enumerates the weekly sales forecast and the monthly sales forecast for 10 different retailers respectively, and uses the two algorithms to make the forecast volume experiment in Python.

Figure 13-14 shows the sales sequence diagram of a retail customer's cigarettes from April 2020 to May 2022, as well as the comparison between the actual sales volume and the predicted sales volume of GCN-LSTM from the beginning of 2022, and the prediction accuracy of the two algorithms is compared. As can be seen from the figure, the average prediction accuracy of shallow neural network is 81%, and that of GCN-LSTM is 89%. It shows that the GCN-LSTM algorithm has a better prediction effect on sales volume, and the predicted value is closer to the real value.

6.4 Experimental Conclusion

In conclusion, the neural network algorithm using feature fusion and the GCN-LSTM hybrid model algorithm can make up for the shortcomings of traditional manual cigarette delivery. GCN-LSTM algorithm is closer to the actual sales in the experiment, and an average accuracy is 89%. It can predict the quantity of cigarette marketing accurately, so as to achieve intelligent and automated cigarette marketing and improve the work efficiency of commercial companies. To solve the problem of intelligent precision marketing of goods, this research proposes a shallow neural network algorithm and a GCN-LSTM hybrid model algorithm, which are used to forecast the marketing volume respectively. Under the experimental scenario of forecasting the weekly sales volume, monthly sales in different time periods of Zhen-long and Nanjing, it is verified that the average prediction accuracy of the shallow neural network is 81%, and the average prediction accuracy of GCN-LSTM is 89%. It can be seen that the prediction results of the GCN-LSTM hybrid model algorithm are more consistent with the actual sales data, indicating that the algorithm can improve work efficiency and better control the marketing of goods to different retailers, so as to achieve precise regulation and achieve intelligent marketing. This research can provide reference value for commercial companies in the field of commodity marketing. In the future, more marketing data will be collected and more effective algorithm models will be designed.

7. Conclusion

In order to solve the problem of intelligent precision marketing of goods, this study proposes shallow neural network algorithm and GCN-LSTM hybrid model algorithm. The two intelligent algorithms are respectively used to predict marketing volume. Under the experimental scenario of predicting the sales of Zhenlong (hard Lingyun) and Nanjing (hard Hong) in a week, a month and sales in different time periods, it is verified that the average prediction accuracy of shallow neural network is 81%, and the average prediction accuracy of GCN-LSTM is 89%. It can be seen that the prediction results of GCN-LSTM hybrid model algorithm are more consistent with the actual sales data, which indicates that the algorithm can improve the work efficiency and better control the product marketing to different retail customers, so as to achieve accurate regulation and realize intelligent marketing. This study can provide reference value for commercial companies in the field of commodity marketing. In the future, more marketing data will be collected and more effective algorithm models will be designed.

Acknowledgement

The authors thank ZHANG Mi-Mi, LIU Shen-Shen and GAO Lian-Xin for providing helpful statistical suggestion.

Reference

[1] Li Xian-zong. Research on product recommendation based on double Knowledge Graph and graph convolutional neural network. Master, Jilin University, 2022.
[2] J. Luo Chang-lin, Chen Meng.Research on guidance training strategy of product marketing based on WEFCMO algorithm. Journal of Hebei Academy of Sciences, 2020, 37 (02), 1-7.
[3] Ke-Miao, Huang Hua-guo. E-commerce commodity sales forecasting method based on LSTM neural network. Journal of Fu-qing University, Fujian Normal University, 2020, (05), 83-89.
[4] Han Ya-Juan. Gao Xin. E-commerce product sales prediction based on Machine learning combination model. Application of Computer Systems 2022, 31 (01), 315-321.
[5] Zheng Yan. Qing dynasty. Xiao Y J. Research on forecasting model of commodity demand based on time series. Journal of Chongqing University of Technology (Natural Science) 2019, 33 (09), 217-222.
[6] Huang Wen-Ming. Research on product sales forecasting based on deep learning. Master, Nanjing University of Science and Technology, 2019.
[7] Huang, D., Liu, H., Bi, T., Yang, Q., GCN-LSTM spatiotemporal-network-based method for post-disturbance frequency prediction of power systems. Global Energy Interconnection 2022, 5 (01), 96-107.
[8] Wu Yu. Research on feature fusion mechanism and application of deep neural networks. Ph.D., Sichuan University, 2021.
[9] Tan Ying-Ying, Jun-li wang, Zhang Chao-bo. A survey of text classification methods based on Graph Convolutional Neural network. Journal of Computer Science, 1-14.
[10] Li Jianhui, Zhi-Hong shen, Meng X F. Scientific Big Data management: Concepts, Technologies and systems. Computer Research and Development 2017, 54 (02),
[11] Gori M , Monfardini G, Scarselli F. A new model for learning in graph domains[C]// IEEE International Joint Conference on Neural Networks. IEEE, 2005.

[12] Li Zhang, Heda Song. Nikolaos Aletras, Haiping Lu. Node-Feature Convolution for Graph Convolutional Networks[J] Pattern Recognition,2022.
[13] Bruna J, Zaremba W, Szlam A, et al. Spectral Networks and Locally Connected Networks on Graphs[J]. Computer Science, 2013.
[14] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering[J]. Advances in neural information processing systems, 2016, 29: 3844-3852.
[15] Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks[C] //2016.
[16] Velikovi P, Cucurull G,Casanova A , et al. Graph Attention Networks[C]// 2017.
[17] Zhao Ling, Song Y, Zhang C, et al. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, PP (99):1-11.
[18] QI Bailin, Guo Kunpeng,Yang Bin, et al. Air quality prediction based on GCN-LSTM [J]. Application of Computer Systems, 2021.Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.
[19] Song Jihong, Wang Tongtong. Encrypted traffic classification method based on feature fusion convolutional neural network. Information and Computer (Theoretical Edition) 2021, 33 (02), 49-51.
[20] Dengyi Huang et al. GCN-LSTM spatiotemporal-network-based method for post-disturbance frequency prediction of power systems [J] Global Energy Interconnection, 2022 (2),96-10.