Fuzzy Systems and Data Mining VIII
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220393

Commodity Width Measurement Based on Big Data

Yan-bing LIU[a], Hao-ran ZHU[a,1], Hao CHEN[a], Zhang-sheng WU[b], Hai-bo WANG[b], Hai-ying LI[b], Ousmane Boubacar-Maiga[c]
[a]*China Tobacco Guangxi Industrial Co., LTD., Nanning 530001, China*
[b]*Zhong ke Know (Beijing) Technology Co., LTD., Beijing 100190, China*
[c]*Anhui Unibest company Ltd., China*

Abstract. Evidences about the impact of commodity width on brand performance remain fragmented. The traditional measurement of commodity category width is characterized by a quantity of manual work and repetition. A fusion model of convolutional neural network (CNN) and long short-term memory (LSTM) is proposed to solve the issue. In order to assure the universality and applicability of the findings, a vast consumer data set covering two retailers and two good categories is used for the measurement. The calculation results show that a composite model has the higher extraction accuracy than any single model on the average, and CNN or LSTM model alone will lead to the lower accuracy and higher error value. Convolutional neural network model possesses of powerful feature extraction, and the accuracy capacity which can be improved by CNN-LSTM fusion model. The mentioned fusion opens a new way for the measurement of commodity width.

Keywords. Convolutional neural network, Commodity category width, Big data, Long short-term memory, Economic field

1 Introduction

Various studies have shown that loyalty to commodity brand is high. Taking cigarette brands for instance, Di Franza, Eddy, Brown, Ryan and Bogojavlensky in 1994 reported that 51% of young people involved in-depth interviewing said that they were still keeping smoking the same brand since they started smoking [1]. Siegel.et al. conducted an extensive survey on American adult smokers and found that only 9% changed their brand each year [2]. In 2000, the research conducted on other consumer goods markets by Ehrenberg has proved the performance of this brand loyalty differentiation in advance [3]. Recently, marketing research and development staff from different tobacco companies have sought to understand consumer preferences through big data by focusing their actual purchases. However, there are obvious regional characteristics such as a strong category planning and vague brand layout should be taken into account.

Chinese market products in general have brand management and maintenance mechanism deficiency [4-6]. Therefore, a large number of scholars show interest in the research on the category planning of product. For example, Xie Ling-Yun [7] made an

[1] Corresponding Author, Hao-ran ZHU, No. 28, Bei-hu South Road, Xi-Xiang-Tang District, Nanning, Guangxi province, E-mail: 736114074@qq.com, 270377937@qq.com

in-depth analysis of the development status of cigarette brands and found that optimizing the brand structure was conducive to enhancing the competitiveness of brands. Yue Yang [8] found that a certain brand had a high market share and marketing power, and the competitiveness of the merchandise brand shows a trend of dynamic change, which is due to the enterprise's ability to build a strong brand and its power to support the development of a strong brand continuously, Teng Jie [9] believed that the change of market position of product brands was affected by the sales and structure of goods brands. Few scholars have made in-depth studies on how to make use of the existing time series data of companies to assist brand planning by rationally regulating the number of merchandise category, maintaining the appropriate number of products, and improving the market scope of superior product.

How to establish a credible product brand structure to identify the number of relevant commodity category at different price in a given region, and establish an organic and a cohesive brand consumer community in order to cultivate merchandise brands and coordinate the marketing activities? Three contributions are as follows on the basis of previous studies. Foremost, it is the first time to measure category width by a deep learning model considering the heterogeneity in the empirical model, and establish the empirical evidence of category width on brand value. Secondly, the features of category width are extracted by the convolutional neural network as the feature set, which mitigated the influence of category width on market share performance. Lastly, the cumulative effect of time is much greater than the effect of the instantaneous moment. It is the direction of brand value that depends on the overall equity level of the category.

2. Research methods

2.1. convolutional neural network (CNN)

Convolutional Neural Networks (CNN) possess a powerful feature extraction ability [10], and it includes input layer, convolutional layer, pooling layer, fully connected layer and output layer [11]. Data generated by connecting the output of the neuron in the prevenient layer with the convolution kernel, each neuron has multiple convolution layers and pooling layers. As is shown in Figure 1. CNN make full use of local connection and shared weights. Firstly, accessing to effective features from the original data by convolution and pooling layers alternately. Secondly, from a special processing of the data layer, the local characteristics of automatic extracting data is achieved to generate valuable information, and then selecting the feature vector through the pooling layer, and the feature dimension reduction process, the fully connected layer serves as a hub, automatically classifies and summarizes the learned features, and finally generates two-dimensional features from the output layer [12].

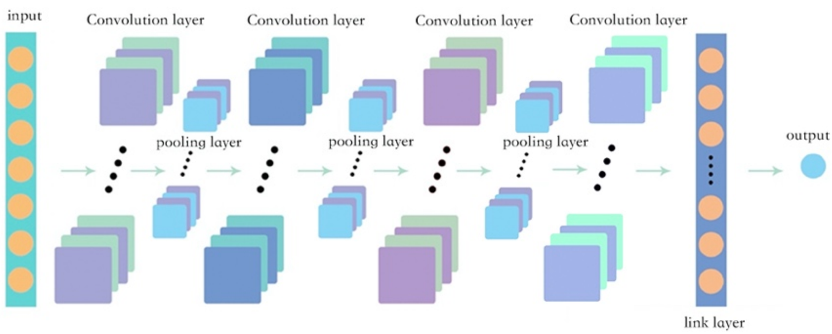

Figure 1 Architecture of convolutional neural network

$$C=f(F * h_{i:i+l-1} + b) \quad (1)$$

As is Formula 1 shown, the convolution kernel is F which belongs to a member of R^(l*d^'). It generates height and width of the convolution kernel parameters, and they are L and d^'. In addition, b is the bias parameters. ReLU is a nonlinear function. Output vector of LSTM hidden layer ranges from the i to the i + l – 1, by the convolution operation, local characteristic vector is obtained [13]. The target is to sample the result of the convolution, reduce the size of the vector convolution to avoid over fitting. Combining all sampling characteristic values as the output of the CNN [14]. ReLU is the activation function [15]. The number of CNN neurons in this study is set to 32, 64, 128, 256, the filter size is 2, and size of the pooling layer is 2.

2.2. Long short-term memory (LSTM)

Though RNN has a powerful dynamic system, practice has proved that this method is problematic with RNN training data, which is prone to gradient disappearance and gradient explosion [16-17]. This means that RNN cannot recall inchoate information [18]. Fortunately, LSTM has solved the problems above, LSTM is a special recurrent neural network [19] including input gate, output gate and forgetting gate , effectively controlling the path of information transmission. That is why LSTM possesses much stronger learning ability than RNN.

Equations (2)-(7) are all formulas needed for calculation at LSTM node:

$$Ft=\sigma\left(w_f * [h_{t-1,}x_t] + b_f\right) \quad (2), \qquad i_t = \sigma \ (w_i * [h_{t-1,}x_t] + b_i \quad (3)$$

$$z_t = tanh \ (w_c * [h_{t-1,}x_t] + b_c \quad (4), \qquad c_t = f_t * c_{t-1} + i_t * z_t \quad (5)$$

$$o_t = \sigma \ (w_o * [h_{t-1,}x_t] + b_o \quad (6), \qquad h_t = o_t * \tanh(c_t) \quad (7)$$

G (t) is invoked as the input unit, H (t) is invoked as the state output unit, M is invoked as a memory unit, I (t) represents the input gate, O (t) represents the output gate, and F (t) represents the forgetting gate. As can be seen from Figure 2, reading, writing and forgetting operations of M are managed by three gating units. Let the output unit of the input time state be as H(t) = O (t) Tanh (M(t)) (6) where σ represents the Sigmoid activation function [20].

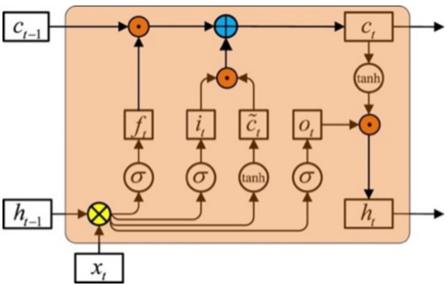

Figure 2 Structure of LSTM model [21]

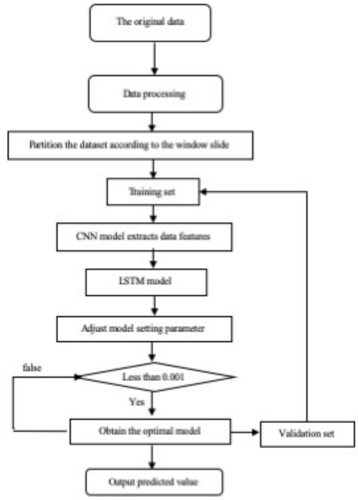

Figure 3 CNN-LSTM's processing steps

The computational state of LSTM is described as the values of forgetting gate, input gate, output gate, candidate state based on the external state at a previous moment and the input at the current moment. In addition, internal states of the previous moment and the last step, the values of forgetting gate, input gate and candidate state are calculated to update the internal state, the information is passed from the current internal state to the external one [22].

A single LSTM neuron is shown in Figure 2, where σ represents the activation function called Sigmoid, Tanh function is applied to change the value size, and the value of output ranges from −1 to 1. The forgetting gate is used for controlling whether the state of the previous moment is retained to the current neuron state to gain memory screening's end. The input gate gives the state value at the previous time and the current input value to activation function Sigmoid as an input [23], so as to get an importance value that determines how up-to-date the information is, and the tanh function is used to process the state value and input information of the previous time to obtain the candidate unit state, the output gate controls the final output of the unit state. The unit state is filtered by the output gate and compressed by the tanh function to obtain the final output of the units [24].

2.3. CNN-LSTM Hybrid Neural Network

The CNN-LSTM hybrid neural network model takes the time series features as the input of the network, the step size is set as 1, and the size of the unit feature map is 16*16. The CNN-LSTM hybrid neural network is divided into three layers. CNN is designed with 4-layer convolution (Conv2D) and the number of convolution kernels is set to 32,64,128 and 256 respectively. CNN can extract many useful features from the input time series data according to the size of the kernel, that is to say, CNN can extract features reflecting a narrower or wider time period of multivariate time series data by setting the kernel size smaller or larger [13]. The first layer: remove the original feature and perform one-hot Encoder to map it into K-dimensional space. K is the encoded vector dimension, and take its index value to obtain the new feature. The second layer: input the new feature into the convolutional neural network, set several convolutional layers and pooling layers, and add the Dropout layer to randomly select some nodes of the hidden layer to prevent overfitting. The relationship between features in the local perceptual domain is obtained by means of convolutional kernel sliding and weight sharing, and high-impact features are extracted. The third layer: the output of CNN is taken as the input of LSTM neural network, and hierarchical LSTM is used for classification to obtain the predicted value of category width [25]. This is shown in Figure 3.

2.4. Evaluation criteria

The root means square error (RMSE) and mean absolute percentage error (MAPE) of common regression models are selected in this paper. The former represents the stability of the model, the lower its value is, and the more stable the model gets. The latter is used to prove the accuracy of the model, the smaller the value is, and the higher the accuracy of the model becomes. They are used to verify the performance of the proposed model in predicting category width.

$$\text{MAE}=\frac{1}{n}\sum_{t=1}^{n}|y-\check{y}_t| \quad (8) \qquad \text{RMSE}=\sqrt[2]{\frac{1}{n}\sum_{t=1}^{n}(y-\check{y}_t)^2} \quad (9)$$

3. Experiment and conclusion

3.1. Data sets

After obtaining the authorization, data for this paper was collected by connecting with the database of the Tobacco Bureau spanning from July in 2019 to January in 2022. A retailer's maximum sales were RMB 609,000, and it was on September 9th in 2019. The minimum sales excluding the mean and standard deviation values were RMB 57,000, and it was on January 26th in 2020. The input data are divided into four categories such as market factors, growth factors, brand positioning factors, and the theoretical coefficients. The missing value is replaced by the mean value. In this paper, convolution kernel automatically extracts data features, and LSTM architecture receives features extracted from the encoder (CNN) as its input layer, simultaneously adjusts data trained in order to discover connections from the input to output sequences.

Differing from other deep learning models such as LSTM and CNN, CNN-LSTM shows none generation vector series. Reading and encoding the input sequence depend on encoder model, reading the encoded input sequence and marking single-step predictions for every factor in the output sequence relay on decoder model, two of parts are the core of the CNN-LSTM Model. CNN Encoder is composed of four convolutional layers and two maximum pooling layers. The first convolutional layer takes charge of reading the input sequence and projecting the result onto the feature map. Repeat the action several times and amplify any salient features until it can remember these features. The kernel size of reading the input sequence is 4 steps, and each convolutional layer has 32, 64, 126 and 256 feature maps. The maximum pooling layer predigests the feature map by the aid of signal value. Feature map extracted is unfolded a long vector to the decoding model as a input by the pooling layer, that is LSTM hidden layers whose parameter is respectively 16, 32 and 64. These layers consist of the decoder model that reads the input sequence and outputs a vector of elements and snatches features from the input sequence. First, the internal representation of the input sequence is copied several times, for each time step in the output sequence. This sequence of vectors will be assigned to the LSTM decoder. The fully connected layer is used to interpret each time step in the output sequence before the final output layer, which means that the same layer will be added for each step in the output sequence.

The CNN and LSTM layers establish its structure to predict cigarette brand width by complex features extracted from multiple sensor variables and complex irregular patterns stored. The architecture of the model should be altered by the patterns and parameters of the component network layers. CNN-LSTM consists of CL, PL, LSTM layers and temporal distribution layers. Each layer manages the number of filters, kernel size, and step size. According to the characteristics of the learning data, these parameters adjusted may make speed and accuracy of the prediction different. The attributes of the input record are vital because they are conductive to changing the parameters and creating the best model for the width of the category. Cigarette marketing data is a multi-source data set consisting of records with 13 variables every day. Each separate subdivision yields a value for valuable forecast interval. Two times distribution layers are used to explain every time step in the output sequence before output layer. The CNN-LSTM hybrid topology is applied to predict the structure of the category width. The convolutional and pooling layer extracts the multivariate variables and inputs variable features, and they are changed to the LSTM layer with noise removal. The CNN-LSTM method can output the category width after a fully connected hierarchy.

3.2. CNN extracts the characteristics of category width

Tobacco commercial company should grasp the development quantity of the cigarette brand specifications to formulate the cigarette brand development plan, and reasonably plan the cigarette brand width, so as to optimize the category structure of the whole market. Industry, market, growth and business operation are the four factors that commercial tobacco companies need to comprehensively consider setting the category breadth.

There is a limit value of the number about single products in the cigarette brand system. The width of the category is helpful to expand more consumer groups. However, the evolution of a brand doesn't depend on the numbers of category, but it depends on the following principles: (1) From the industry level, in accordance with the brand layout policy of stabilizing the first class, raising the second class, increasing the third class, reducing the fourth class, and controlling the fifth class, implemented the national Bureau of tobacco category control. First class, second class, third class, fourth class and fifth class cigarette input experience coefficient, so as to ensure that the formulated category structure tends to the direction of development, (2) From the market level, it can meet consumers' consumption habits, consumption attitudes and consumption characteristics, and reflect the market environment realistically, which highlights both the rationality of category formulation and also follow the nature of market demand, (3) From the perspective of growth, starting from the potential growth and changing characteristics of the category can ensure that the formulated category structure respect the nature of the market as well as conforming to the change, (4) From the perspective of positioning, comprehensive consider consumers, enterprises and the characteristics of the category. CNN model is used to extract the features of category width such as theoretical coefficient, sale volume, sales, gross profit, growth rate of sales, growth rate of gross profit, category role.

3.3. Operation procedure of cigarette category width

(1) Set the theoretical coefficient

(2) Calculation of market factors

The annual sales, annual sales, gross profit and average sales of each category were statistically analyzed, and the average characteristics are extracted. After the statistics are completed, the forced distribution is carried out to calculate the data score of each category.

(3) Calculate growth factors

The annual sales growth rate, sales growth rate and gross profit growth rate of each category are statistically analyzed, and the average value characteristics are extracted. After the statistics are completed, the forced distribution is carried out to calculate the data score of each category.

(4) Calculation of positioning factors

The role positioning matrix of category is constructed, the market value and market growth rate of category are statistically analyzed, the role of category is calculated, and the category is divided into key category, conventional category, growing category and general category, so as to determine the distribution of category role.

(5) Calculate the comprehensive coefficient of the width of each category

According to the formula: comprehensive coefficient of category width = sales score + gross profit score + sales growth rate score + gross profit growth rate score + category role score.

(6) Calculate the width of each category

According to the formula: the reference value of category width planning = total single-product quota * comprehensive coefficient of category width, multiply the comprehensive coefficient of each category width by the total planning value of annual specifications (total single-product quota), and round up to get the reference value of annual category width planning (excluding imported cigarettes and cigars).

(7) Output of category width planning results

Combining with the annual sales scale, we determine the total number of cigarette specifications on the basis of the approved quantity. The limit value of floating ratio doesn't exceed 10%, we adjust the width of the category and the floating ratio, rounded to get the annual width range of each category.

3.4. Data preprocessing

The preprocessing in this paper includes beneficiating missing value, and preconditioning outlier and data standardization.

(1) Through the analysis of the original data, the sliding window method is used to serialize the original data, and the Z-score method is used to standardize the original data.

(2) Different from the single model, a CNN-LSTM hybrid neural network is proposed to improve the accuracy and stability of brand prediction width. The CNN sub-module is used to explore the correlation and nonlinear characteristics of data at adjacent time points, while LSTM captures the long-short-time correlation of time series data [22].

3.5. CNN-LSTM Category width measurement model prediction algorithm

Input: theoretical coefficient, market factor, growth factor and positioning factor.

Output: the planning reference value of the category width,

Algorithm process:

1: read the date = {}, the input data is read and the category width features are constructed.

2: reate x= {}, construct category width characteristics.

3: C = ReLU (Convolution (32,16,16)), the feature vector of category width was constructed, and the parameters of the model were set. The size of the convolution kernel was 16*16, and the number was 32.

4. Dimensionality reduction of feature vectors, set the number and size of pooling layers,

Co = Max pooling 16 (dec)

5. Train the LSTM model to get a new vector matrix.

L = LSTM (output = 256, activation = 'tanh')

6. class=Softmax () category width classification.

3.6. Experimental results

The architecture provides the results for CNN-LSTM, whose performances of classification are decided by the number of layers and units of each layer. As a parallel feature extractor, CNN automatically extracts features from element columns combined. LSTM is a recurrent neural network that has been grown in performance and structure, which is to learn complex nonlinear data sets and extract advanced features automatically, and integrated structural elements of deep learning structure can improve its function. However, the CNN layer is excluded, its performance gradually decreases. Adding more convolutional layers to balance complexity, because performance degradation appears on the system. The performance of the proposed model decreases by 1% with the increase of the structure complexity of the model, the optimal selected parameter is received from optimizing the hyperparameters in a feasible independent scope in the parameter distance. The algorithm set time step and size those are respectively 1 and 7, LSTM units are 10 and the epoch selected 90.

4.The differences among model structures

4.1. Contrast to other machine learning models

Five different machine learning models are used to verify the efficiency of the CNN-LSTM model. parameters of various machine learning models are shown in Table 1. In the process of calculation, we change parameters to obtain the best performance of the proposed model. At last, the proposed CNN-LSTM model have the lowest mean absolute error (MAE) compared with (KNN, SVR, LR, DTR, LASSO). Experiencing linear regression analysis, the machine learning model is ranked as LASSO, DTR, SVR, KNN, LR. The results of CNN-LSTM method proposed in this paper are compared with the linear regression model. The CNN-LSTM method performs best among these models in predicting the global characteristics of category planning. It is benefit to predict global or local features. However, linear regression models are not fit to predict the local and global characteristics of this data set. CNN-LSTM is more well in processing time series. CNN and LSTM model improves the accurate prediction of category width output.

Table 1. Parameter Settings of machine learning model description

Model	Model Model Description
1 KNN	neighbors=10
2 LR	Number up=True, normalize=false
3 DTR	Max-depth=none, min-samples-split=1
4 LASSO	Alrha=0.1
5 SVR	Kernel= ‘rbf’ ,gamma= ‘scale’, c=1.0, epsilon=0.2

The RMSE and MAE error values are used to compare the performance of different machine learning models. Five machine learning models are worse than CNN-LSTM model in performance, whose error values close to each other in contrast to other models, RMSE is 13.15 and MAE error value is 11.6 respectively from KNN with the lowest skills. As is shown, the developed architecture lows the error value by 50%, resulting in prediction accuracy. These parameters are shown in Table 2.

Table 2. Setting of single model parameters

Model	Model parameters
CNN-LSTM	The convolution kernel has three layers, the size is 16*16, the number of LSTM hidden layers is 32, and the CNN has four layers
LSTM	The number of fully connected neurons was 30, the number of LSTM hidden layers was 16, and the optimizer was Adam
SVR	The number of neurons is 32. Activation function, RELU
CNN	There are 32 connection layers, three filtering layers, the activation function is RELU, and the optimizer is Adam.

4.2 Performance comparison of a single deep learning model

The CNN-LSTM hybrid model perform best in predicting category width among the classical LSTM methods. The MAE of the CNN model is 5.26, which is the adjacent to algorithm to the CNN-LSTM model. LSTM performed least and MAE value is 9.55. CNN or LSTM model alone will lead to a lower accuracy and higher error value. Giving full play to core strengths of a single hybrid model, it will stand out of the rest. LSTM requires more hidden layers to predict more precisely than CNN. Limited by research, the above contents will not be discussed.

CNN-LSTM convolution kernel has three layers, the size is 16*16, the number of hidden layers of LSTM is 32, and CNN has four layers.

The number of fully connected neurons in LSTM was 30, the number of hidden layers in LSTM is 16, and the optimizer is Adam.

The number of SVR neurons is 32, activation function, ReLU.

CNN has 32 connection layers, three filtering layers, the activation function is RELU, and the optimizer is Adam.

4.3. Differences in prediction performance according to time resolution

The CNN model, LSTM model and CNN-LSTM model are compared with the actual prediction results. The parameter of Look back is 10 steps, and look forward is 1 step ,2 steps and 3 steps (as shown in Figure 4). The CNN-LSTM model are superior to the LSTM model in predicting local and global features, LSTM outliers evidently decrease by time. The CNN model can approach to performance of CNN-LSTM in predicting local and global features, which is better than the another one.

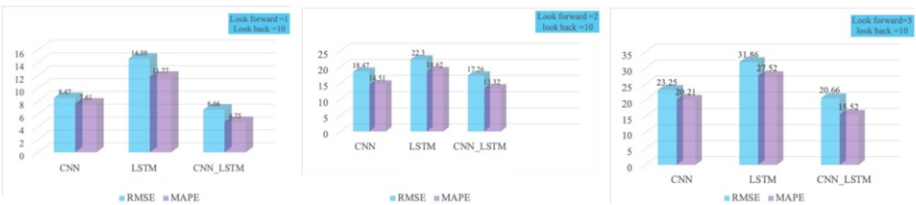

Figure 4 Prediction performance from 1 step to 3 steps

4.4. Discussion on prediction model of CNN-LSTM network

For multi-step prediction, a comparison shows a considerable difference among these machine learning models. The CNN-LSTM architecture has a lower mean MAPE than LSTM for all retrospective and forward time measures. On the one hand, the percentage enhancement of MAPE seems to decrease slightly when the time increases from 10 to 14 steps. With the forward time step increasing, the separation between the MAPE value of proposed model and the LSTM model becomes larger. The CNN-LSTM model is promoted by 5.65%, in which the review step is 10 times and the outlook step is 1 time. However, the proposed model tends to expand this value as the forward time step increases. From 7 steps forward and 14 steps back, it can be seen that MAPE increased by 7.13%. Therefore, the hybrid model proposed in this paper outperforms the LSTM model in both single-step and multi-step forecasting. A 14-day review is more effective than a 10-day review in predicting a 7-day scenario. This is due to the longer prediction time and the need more times to improve accurate predictions. In terms of training time, the proposed architecture has much longer training time than LSTM model. We prefer to use the model with the lowest training time and some compromise for accuracy, because the impact of category width on brand value is a long-term cumulative process.

For less resources and the imbalance of the distribution of waste, further strengthen brand on the transverse specification number. According to the evaluation of the brand value, set the clear goal, quit brand specification list, statistics of the sales scale of all cigarette brands in the given market as a reference, the annual growth in sales, integrated determine the total number of reasonable cigarette specifications.

5. Conclusion

In this present research, CNN-LSTM was used to construct a comprehensive category width measurement model, moreover a certain number of dominant specifications for each price level were constructed to provide a basis for the different industry to select the brand composition in different prices to set up the overall brand layout planning. By setting the number of hidden layers of different CNN-LSTM models, the research can determine the best parameters of the model. What is more the deep learning algorithm is applied to predict the category width for the first time in domain of such research. Taking the tobacco industry data as an example, the effectiveness of the model is verified, and the superiority of the performance model is proved by comparing the single models with CNN-LSTM. Most importantly, it is the first time that deep learning model has been used to measure the category width. The empirical evidence of the effect of category width on brand value was established by considering the heterogeneity in the empirical model. By Calculating each category in the brand comprehensively, according to the results of the brand evaluation division brand specification in the role of category, the research makes to clear the value of each category scores, measure the scientific and rational commodity combination structure, finally confirm the dominant product rules. This model safeguard rules such as potential rules, general rules and the number of new rules as well as brand to form the result of the category and reasonable layout. In addition, the model used can allow scholars and companies to define the market position of each brand specification in the category and Optimize the category layout to form a continuous improvement of brand layout planning and management closed-loop continuously. The model constructed in this study has good application value and has certain reference significance for the brand planning of tobacco companies.

We arrive at a conclusion that consumers' long-term preference or lack of preference for a particular product can have an impact on category breadth and profitability, which can also provide opportunities for competitors to react and attack brand value.

References

[1] DiFranza, J. R., Eddy, J. J., Brown, L. F., Ryan, J. L., & Bogojavlensky, A. (1994). Tobacco ac- quisition and cigarette brand selection among youth. Tobacco Control, 3(4): 334.

[2] Siegel, M., Nelson, D. E., Peddicord, J. P., Merritt, R. K., Giovino, G. A., & Eriksen, M. P. (1996). The extent of cigarette brand and company switching: results from the Adult Use-of-Tobacco Survey. American Journal of Preventive Medicine, 12(1):14–16.

[3] Ehrenberg, A. S.C. (2000). Repeat-buying: Facts, theory and applications. Journal of Empirical Generalisations in Marketing Science, Feb 5:392–770.

[4] NIU Yanjuan. Research on brand Marketing of cigarette Commercial Enterprises. Gansu: Lanzhou University,2012.

[5] YANG Qi. Reflections on the construction of cigarette category. Eastern Corporate Culture, Jan (03):156.

[6] LI Xuhua. Discussion on the construction of cigarette category. China Business & Trade, March (30):239-240.

[7] XIE Lingyun. Construction of cigarette brand planning model based on data-driven. Strait Science, (08):111-116.

[8] YUE Y.Comprehensive evaluation model of key cigarette brand competitiveness based on multiple evaluation indexes . Modern Business,2010(14):86-87.

[9] TENG Jie. Research on cigarette brand cultivation strategy under the New Normal. Modern Commerce & Trade Industry, 201,42(01):65-67.

[10] A. Krizhevsky, I.Sutskever,G.E.Hinton Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems,2012:1097-1105.

[11] Ali Agga, Ahmed Abbou, Moussa Labbadi, Yassine El Houm,Imane Hammou Ou Ali. "CNN-LSTM efficient hybrid deep learning architecture for predicting short- term photovoltaic power production", Electric Power Systems Research, 2022
[12] XIONG-Hao, LIU Jia-yong, Wang Jun-feng. Network Traffic Prediction Based on Neural Network and Autoregressive Model computer Applications, 201,41(S1):180-184.
[13] Jianfeng Deng, Lianglun Cheng, Zhuowei Wang. "Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification" Computer Speech & Language, 2021
[14] Yan-hui Liang, Yu-bing, Qin-lu. Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. Expert systems with application.2022(6):1-39.
[15] A.Maas,A.Hannun,A.NgRectifier nonlinearities improve neural network acoustic model. International Conference on Machine Learning (ICML), in ICML Workshop on Deep Learning for Audio, Speech and Language Processing (2013)
[16] Pascanu R., Mikolov T., Bengio Y.On the difficulty of training recurrent neural networks. International Conference on Machine Learning, PMLR (2013):1310-1318.
[17] Bengio Y., Simard P., Frasconi P.Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw., 5 (2) (1994):157-166.
[18] Canizo M., Triguero I., Conde A., Onieva E.Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study .Neuro computing, 363 (2019):246-260.
[19] HochreiterS., SchmidhuberJ.Long short-term memory.Neural Computer, 1997(3):1735-1780.
[20] ZHANG Peng, Yang Tao, LIU Yannan, FAN Zhiyong, DUAN Zhaobin. Feature extraction and prediction of QAR data based on cnn-lstm . Application research of computers, 2019,36(10):2958-2961. 2961.
[21] ZhipengShen, XuechunFan,LiangyuZhang,HaomiaoYu.Wind speed prediction of unmanned sailboat based on CNN and LSTM hybrid neural network. Ocean Engineering.2022(254):53-60.
[22] Zhipeng Shen, Xuechun Fan, Liangyu Zhang, Haomiao Yu. "Wind speed prediction of unmanned sailboat based on CNN and LSTM hybrid neural network" Ocean Engineering, 2022
[23] Hong Xu, Tao Tang, Baorui Zhang, Yuechan Liu. "Automatic sentiment analysis of public opinion on nuclear energy", Kerntechnik, 2022
[24] LU Jixiang, ZHANG Qi-pei, YANG Zhihong, TU Mengfu, LU Jinjun, et al. Short-term Load Forecasting Method Based on CNN-LSTM Hybrid Neural Network Model Automation of electric power systems.2019,43(8):131-137.
[25] XING Yang, WANG Shaopeng advertising Click-through Rate Prediction Model Based on Convolution-LSTM Network . Computer engineering and applications, 2019,55 (2): 193-197.