2020 5th International Conference on Computer and Communication Systems

# A Commodity Review Sentiment Analysis Based on BERT-CNN Model

Junchao Dong
Guangxi Key Laboratory of Trusted Software
Guilin University of Electronic Technology
Guilin, China
e-mail: 764919430@qq.com

Feijuan He *
Department of Computer Science
Xi'an Jiaotong University City College
Xi'an, China
e-mail: hfj@xjtu.edu.cn
*Corresponding authors

Yunchuan Guo
Institute of Information Engineering
Chinese Academy of Sciences
Beijing, China
e-mail: guoyunchuan@iie.ac.cn

Huibing Zhang
Guangxi Key Laboratory of Trusted Software
Guilin University of Electronic Technology
Guilin, China
e-mail: zhanghuibing@guet.edu.cn

***Abstract*—Sentiment analysis through the investigation on commodity reviews will be of great importance to commodity quality improvement of the seller and subsequent consumption choice of buyers. The accuracy of the existing sentiment analysis models remains to be further improved, so a BERT-CNN sentiment analysis model, an improvement of the original BERT model, was proposed in this paper in order to improve the accuracy of commodity sentiment analysis. Firstly, BERT model was constructed, and then a representation layer was input into the model to encode the review texts; after then, CNN semantic extraction layer was utilized to extract local features of the review text vectors, BERT semantic extraction layer to extract global features of the review text vectors and semantic connection layer to fuse features extracted by the two complementary models; in the end, a sentiment analysis of online commodity reviews was performed via the sentiment classification layer. The experimental results indicated that in comparison with BERT and CNN models, F1 value of BERT-CNN model was elevated by about 14.4% and 17.4%, respectively.**

***Keywords-e-commerce; commodity review; review text; sentiment analysis; BERT-CNN model***

## I. INTRODUCTION

In the wake of the development of internet technology and ever-increasing quantity of commodities on e-commerce platforms with uneven commodity qualities, commodity reviews are especially important as the most intuitive embodiment of commodity quality. The commodity review sentiment analysis on e-commerce platforms can help new buyers to select high-quality commodities, sellers to improve commodity qualities and e-commerce platforms to promote outstanding buyers and timely eliminate inferior commodities buyers are dissatisfied with, so as to further improve the competitiveness of e-commerce platforms.

The existing commodity review sentiment analysis is mainly divided into two types: machine learning-based method [1], [2], and deep learning-based method[4], [5]. Literature [3] evaluated service qualities of commercial tenants through a sentiment analysis of online reviews, classified them using a Bayesian model and dug advantages and disadvantages of service qualities of the commercial tenants. Machine learning-based method mainly refers to commodity review sentiment analysis by importing related features established after data processing into the machine learning model, and performance of this method depends a lot on feature engineering. However, commodity evaluation data scale is becoming larger and larger with a growing number of users on e-commerce platforms, so time and manpower needed by feature engineering also abruptly increases. In recent years, Convolutional Neural Networks(CNN) and Recurrent Neural Networks (RNN)-represented deep learning-based sentiment analysis models have been extensively applied to text sentiment classification and sentiments in commodity reviews, etc. by virtue of automatic data feature extraction and favorable experimental results. Literature [6] proposed a sentiment classification model based on gating recursive unit and attention mechanism and classified positive and negative reviews to understand user's opinions over some commodities or services. Nevertheless, the accuracy can be very low only by relying on CNN model while neglecting semantic relations existing in the context in the review texts. Long Short-Term Memory (LSTM) model can't accurately dig semantic relations between subclauses in the review texts, so it can't adapt to the realistic demand for sentiment analysis of long-text commodity reviews containing multiple subclauses.

Therefore, a BERT-CNN based sentiment analysis model of commodity reviews was raised in this paper. This model utilized deep learning method to automatically extract features from commodity review texts, thus solving the artificial participation problem in dictionary construction and feature engineering in the traditional sentiment analysis model of commodity reviews. CNN and Bidirectional Encoder

978-1-7281-6136-5/20/$31.00 ©2020 IEEE

Representation from Transformers (BERT) semantic extraction layers were used to extract local and global features of review text vectors, respectively, and features extracted by the above two complementary models were integrated to solve the problem of single CNN model, naming neglecting semantic relations in the context of review texts, thus effectively mining semantic relations between short sentences in long texts. The experiment was performed on mobile phone review dataset in JD Mall, and the results indicated that the proposed model had higher accuracy than BERT and CNN models.

## II. SENTIMENT ANALYSIS MODEL OF ONLINE COMMODITY REVIEWS

Fig. 1 shows BERT-CNN commodity review sentiment prediction model framework. This model adds an independent CNN semantic extraction layer after a representation layer is embedded into the original BERT model, so as to extract local semantic information in the review texts, and then connects hidden state output by CNN pooling layer with the final hidden state of the first input word block ([CLS]) in BERT model as the fully connected layer, which is then input into the sentiment classification layer for commodity review sentiment analysis [7].

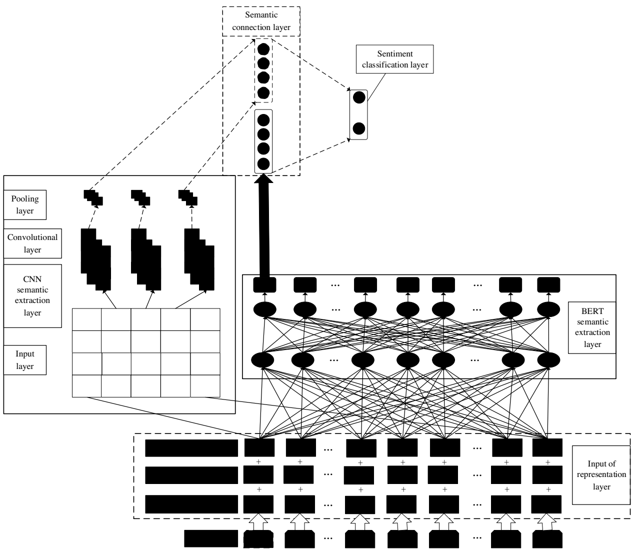


Figure 1. BERT-CNN commodity review sentiment analysis model framework.

### A. Input of Representation Layer

Before the commodity review sentiment analysis, review text features should be summarized firstly so as to standardize the data input into the representation layer. Based on the statistics of commodity review data in three e-commerce platforms—Taobao, Tmall and JD, it's found that commodity review data are mingled with ambiguous figures, emotion icons and English words together with a large quantity of URLs and free chat-like reviews irrelated to commodity sentiment analysis. Therefore, before the data input, they are preprocessed through the following steps.

- A large number of repeated and meaningless free chat-like reviews and URL review data irrelated to the expression of feelings are excluded.
- Fuzzy emotion icons, figures and English words contained in the commodity reviews are replaced with Chinese words having the same meanings.

Different from the encoding using term vectors like word2vec and ELMo in the representation layer of the traditional natural language model, BERT model is a pretraining model which uses sentence vector-level encoding related to the context [7], so it can accurately quantify meanings of the same word in different contexts within the short commodity review text, and meanwhile, it can realize encoding of relations between sentences in the long commodity review text, and the concrete flow is as follows:

*1)* There are numerous Chinese words with abundant meanings. The traditional encoding method taking terms as the unit has a complicated structure and this is adverse to the subsequent semantic extraction by the model, so WordPiece embedding model was used in this paper to code single Chinese words in the commodity review texts, thus greatly reducing encoding scale and complexity of the review texts.

*2)* According to the statistics, the proportion of commodity review texts with number of Chinese words smaller than 300 reaches over 90%, so maximum sequence length is 300 Chinese words in the word block embedding training process. According to the difference of sentence structures of review texts, review texts are largely divided into short phrases and long phrases, and the concrete training process is as follows:

- The language model is constructed using mask LM method for short review texts, the main idea is to randomly mask or replace any Chinese character in the review text, and then the model will predict the masked or replaced part by understanding the contextual content, and the concrete operation method is: 15% token in a review text is randomly replaced and substitute mode includes: 80% probability to be replaced by [MASK], e.g. cellphone quality is very good—> cellphone quality is very [MASK]; 10% probability to be replaced with another token, e.g. cellphone quality is very good—> cellphone quality is very pretty; 10% probability that the original content is kept unchanged, e.g. cellphone quality is very good—> cellphone quality is very good.
- For long review texts, some special tokens [SEP], which are used to judge start-stop positions of the last and the following sentences, are added to semantic logic of the review text as shown in the input in Fig. 2. Review texts with and without context relations are input as input representation layers by 1:1 proportion so that the model can understand subclause relations in the commodity review texts, for instance:

  Input=[CLS]手机 [MASK]非常好 [SEP]我[MASK]喜欢 [SEP]

  Label=have context relation

Input=[CLS]手机 [MASK]非常好 [SEP]买[MASK]面包 [SEP]
Label=no context relation

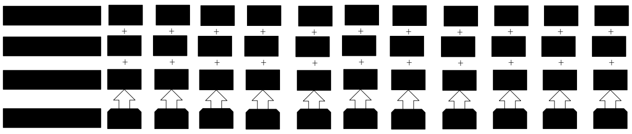

Figure 2. Input representation layer.

Besides input and token embedding, the input representation layer needs to randomly initialize a trainable segmental embedding as shown in Fig. 2, where with the segmental embedding information, the model can judge start-stop positions of the last and the following sentences so as to separate them as shown in Table I.

TABLE I. EXAMPLE OF SEGMENTAL EMBEDDING

| Token embedding | [CLS] | 手 | 机 | 非 | 常 | 好 | [SEP] | 我 | 喜 | 欢 | [SEP] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Segmental embedding | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |

The final output of the input representation layer includes token embedding, segmental embedding and position embedding.

*B. BERT Semantic Extraction Layer*

As shown in Fig. 1, BERT semantic extraction layer is a multilayer two-way decoder with Transformer encoder as basic unit, and the output hidden layer is namely mathematical expression of the commodity review text, where any hidden unit at the hidden layer contains all sentiment information in the review text after passing through the attention mechanism in Transformer encoder. As shown in Fig. 3, Transformer encoder includes four parts: word vector and position encoding, attention mechanism, residual connection, and layer standardization and feedforward [8].

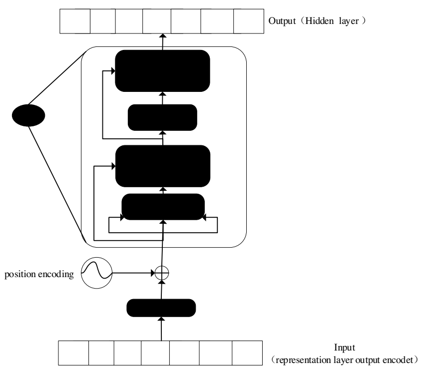


Figure 3. Transformer encoder structure.

*1)* Word vector and position encoding

Position encoding provides position information of each word in the commodity review text for Transformer to recognize its dependency and time sequence characteristics in the review text.

$$X = E(X) + P, X \in \mathbb{R}^{\text{batch-size*seq.len*embed.dim}} \tag{1}$$

Where $E(\bullet)$ is embedding expression corresponding to the word; $P$ is position encoding solved through formula 2-3; batch-size is input quantity of commodity review text; seq.len is length of each commodity review text; embed.dim is embedding dimensionality of each word in the commodity review text.

Position encoding is obtained through linear transformation of equations 2-3,

$$P_{(pos,2i)} = \sin(pos / 1000^{2i/d_{\text{model}}}) \tag{2}$$

$$P_{(pos,2i+1)} = \sin(pos / 1000^{2i/d_{\text{model}}}) \tag{3}$$

Where pos is position of the word in the commodity review; i is corresponding vector dimensionality.

*2)* Attention mechanism

Attention mechanism ensures that each word vector in each review text contains information of all word vectors in this commodity review text. The calculation formulas are shown in 4-5.

$$\begin{aligned} Q &= Linear(X) = XW_Q \\ K &= Linear(X) = XW_K \\ V &= Linear(X) = XW_V \end{aligned} \tag{4}$$

Where $Linear(\bullet)$ is linear mapping; $X$ is commodity review text vector; $W_Q$, $W_K$, $W_V$ are weights.

$$\begin{aligned} X_{\text{attention}} &= SelfAttention(Q,K,V) \\ &= \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \end{aligned} \tag{5}$$

$\sqrt{d_k}$ in the above equation transforms attention mechanism into standard normal distribution; $\text{softmax}(\bullet)$ is normalization which makes sum of attention weights of each word and other words in the commodity review text.

*3) Layer standardization and residual connection*

According to equation 6, layer standardization accelerates model training for standard normal distribution through the hidden layer in the normalized neural network, so as to accelerate model convergence. Residual connection can solve gradient vanishing and network degradation problems as shown in equation 9.

$$LayerNorm(x) = \alpha \odot \frac{x_{ij} - \mu_i}{\sqrt{\sigma_i^2 + \varepsilon}} + \beta \tag{6}$$

Where $\mu_i$ is mean value of matrix row solved by equation 7; $\sigma_i^2$ is variance of matrix row solved by equation 8; $\odot$ is multiplication of elements in the matrix; $\alpha$ and $\beta$ are model training parameters; $\varepsilon$ is used to prevent denominator from being 0.

$$\mu_i = \frac{1}{m}\sum_{i=1}^{m} x_{ij} \tag{7}$$

$$\sigma_i^2 = \frac{1}{m}\sum_{i=1}^{m}(x_{ij} - \mu_i) \quad (8)$$

$$X = X_{embedding} + attention(Q, K, V)$$
$$X + subLayer(X) \quad (9)$$

Where $subLayer(\bullet)$ is implementation function of the sublayer itself.

*4)* Feedforward

Feedforward is formed by linear mapping of two layers and activated by ReLU activation function [7]. The output commodity review text passes through the hidden sequence after processing of input encoding layer and attention mechanism, and the calculation formula is as below:

$$x_{hidden} = \mathrm{ReLU}(\mathrm{Linear}(\mathrm{Linear}(X))) \quad (10)$$

### *C. CNN Semantic Extraction Layer*

CNN semantic extraction layer consists of convolutional layer and pooling layer. Mainly extracting sentiment information of local vectors in the review text, it is a beneficial supplementation for global variable sentiment information at the DERT semantic extraction layer.

*1)* Convolutional layer

The input layer receives review text vector $x_i^0 = \{x_1, x_2, \cdots, x_{300}\}$ output by the input representation layer and uses convolutional layer to carry out convolution operation, so as to accurately obtain local sentiment information contained in the review text, and the output of the convolutional layer is as shown in equation 11.

$$y_{ij}^I = \sigma\left(\sum_{m=1}^{M} w_{m,j}^I x_{i+m-1,j}^{I-1} + b_j^I\right) \quad (11)$$

Where $y_{ij}^I$ is calculated through output vector $x_{ij}^I$ at the input representation layer; $b_j^I$ is bias mapped by the feature j; $w$ is weight of convolution kernel; m is index value of filter; $\sigma$ is ReLU activation function.

*2)* Pooling layer

After the convolution operation, local sentiment information in the review text are transferred to the pooling layer in order to further reduce quantity of parameters, data dimensionality and overfitting as shown in equation 12.

$$x_{hidden} = \max_{r \in R}(y_{i \times T + r, j}^{I-1}) \quad (12)$$

Where T is step length of pooling stride; R is pooling size.

### *D. Semantic Connection Layer*

The main effect of semantic connection layer is to connect output of BERT semantic extraction layer and that of CNN semantic extraction layer and construct overall semantic information vector of the commodity review text as shown in equation 13.

$$S = [S_b, S_c] = (r_b + r_c) \times c \quad (13)$$

Where $S$ is overall semantic information vector; $S_b$ and $S_c$ are semantic information vectors output by BERT and CNN semantic extraction layers, respectively; $r_b$ and $r_c$ are vector row numbers of $S_b$ and $S_c$, respectively; $c$ is vector column number of $S_b$ and $S_c$.

### *E. Sentiment Classification Layer*

The main task of the sentiment calculation layer is to construct a sentiment classifier, acquire score vector of semantic expression $S$ of the commodity review text relative to sentiment label and output the final sentiment label. Sigmoid function is used to do corresponding calculations of output at the semantic connection layer, so as to perform sentiment classification of the commodity review text. The specific formula is as follow:

$$y = sigmoid(S), y \in (0,1) \quad (14)$$

Entropy loss function is used to measure the error between real probability distribution $y$ and predicted probability distribution $y$ of the sentiment label,

$$Loss = -\sum_{r \in R}[y(r) \cdot \log(y(r)) + (1-y(r)) \cdot \log(1-y(r))] \quad (15)$$

Where R is training dataset of the commodity review texts; $r$ is a commodity review text in the training dataset; $y$ is predicted sentiment value of the sample; $y$ is real sentiment value. Parameter training and optimization are carried out through the backpropagation method.

## III. Experiment and Analysis

### *A. Dataset*

The pretraining model of the original BERT model is obtained through training on Chinese Wikipedia corpus. A total of 5,000 positive and 5,000 negative cellphone review data on JD Mall are used in the downstream task of online review sentiment analysis, and data format is shown in Fig. 4. The BERT-CNN sentiment analysis model is trained and tested, and training sets and test sets are divided by the proportion of 7:3.

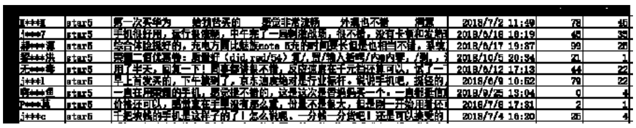

Figure 4. Cellphone review data format from JD Mall.

### *B. Evaluation Indicators*

Precision P, recall rate R and F1 value are used to evaluate the sentiment analysis model. According to real types of the examples and combination of BERT-CNN predicted types, the examples are divided into true positive (TP), false positive (FP), true negative (TN) and false negative (FN) types, and precision P, recall rate R and F1 value are obtained after digitalization, and their computational formulas are respectively:

$$P = \frac{TP}{TP + FP} \tag{16}$$

$$R = \frac{TP}{TP + FN} \tag{17}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{18}$$

### *C. RP-BERT Model Verification*

BERT-CNN model is compared with the original CNN model and BERT model. In order to ensure accuracy and objectivity of experimental results, the three models are put under operation for 10 times on the same training set and test set, respectively, to solve precision, recall rate and F1 value as the final model results as shown in Table II.

TABLE II. COMPARISON OF PRECISION, RECALL RATE AND F1 VALUE OF THE THREE SENTIMENT ANALYSIS MODELS OF COMMODITY REVIEWS

| **Model type** | **Training dataset** | | |
|---|---|---|---|
| | ***Precision*** | ***Recall rate*** | ***F1 value*** |
| CNN | 0.706 | 0.715 | 0.710 |
| BERT | 0.802 | 0.823 | 0.812 |
| BERT-CNN | 0.856 | 0.847 | 0.851 |
| | **Test dataset** | | |
| CNN | 0.702 | 0.709 | 0.709 |
| BERT | 0.798 | 0.820 | 0.820 |
| BERT-CNN | 0.854 | 0.843 | 0.843 |

As shown in Table II, F1 values of both BERT-CNN model and BERT model are higher than those of CNN model in the training set and test set, mainly because the two models are unsupervised pretraining models for large-scale corpuses not needing artificial participation, and their attention mechanism can accurately connect semantic contexts for mining accurate sentence meanings of commodity review texts and semantic logic between subphrases in each review text. In addition, the difference between F1 value of BERT-CNN model in the training set and that in the test set is very small, so the model has favorable generalization ability.

## IV. CONCLUSION

A BERT-CNN commodity review sentiment analysis model was proposed in this paper. CNN was used to extract local features of review text vectors and BERT to parallelization was used to extract global features of review text vectors, and features extracted by the two complementary models were fused to solve the problem of a single CNN model, namely neglecting contextual semantic relations in the review text and also effectively dig semantic relations between short sentences in long texts. The evaluation index, that is F1 value, was used to verify prediction accuracy of the model in real cellphone review dataset on JD Mall. The experimental results showed that compared with the original BERT and CNN models, BERT-CNN model effectively elevated F1 value predicted by the model.

However, the proposed sentiment analysis model only relies on information in the commodity review texts of the users, and in reality, the method which can embody user emotions more intuitively than text information is commodity star level given by the users. Subsequently, a commodity review sentiment analysis model, which integrates commodity review texts and commodity star level will be constructed so that the model can be more accurate and pragmatic.

## ACKNOWLEDGMENT

This research is supported by the National Key Research and Development Program of China (Grant No. 2017YFB0803001); National Natural Science Foundation ( 61662013, U1501252, U1711263, U1811264,61662015); Guangxi Innovation-Driven Development Project (Science and Technology Major Project)(AA17202024); Guangxi Natural Science Foundation (2017GXNSFAA198372); Beijing Information Technology University Fund Project (5221910933);The Teacher Growth Fund of the Education Development Foundation of Guangxi Normal University (EDF2015005);The Funds of Graduate student innovation program Guilin University of Electronic Technology (2019YCXS045).

## REFERENCES

[1] I. Yulietha, S. Faraby, and W. Widyaningtyas, “An implementation of support vector machine on sentiment classification of movie reviews,” Journal of Physics: Conference Series, vol.971, no. 1, p. 012056, 2018.

[2] V. Tama, Y. Sibaroni, “Labeling Analysis in the Classification of Product Review Sentiments by using Multinomial Naive Bayes Algorithm,” Journal of Physics: Conference Series, vol.1192, no. 1, p. 012036, 2019.

[3] P. Sari, D. Alamsyah, S. Wibowo, and D. Adzkiya, “Measuring e-Commerce service quality from online customer review using sentiment analysis,” Journal of Physics: Conference Series, vol.971, no. 1, p. 012053, 2018.

[4] S. Wang, G. Lv, S. Mazumder, and G. Fei, “Lifelong Learning Memory Networks for Aspect Sentiment Classification,” IEEE International Conference on Big Data (Big Data), 2018.

[5] Q. Li, S. Li, S. Zhang and J. Hu, “Tourism Review Sentiment Classification Using a Bidirectional Recurrent Neural Network with an Attention Mechanism and Topic-Enriched Word Vectors,” Sustainability, vol. 10, no. 9, pp. 3313, 2018.

[6] H. Poon, W. Yap, S. Wibowo, and Y. Tee, “Document level polarity classification with attention gated recurrent unit,” International Conference on Information Networking (ICOIN), 2018.

[7] J. Devlin, M. Chang, W. M and K. Lee, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint, vol. 10, no. 04805, 2018.

[8] A. Vaswani, N. Shazeer, N. Parmar, and N. Uszkoreit, “Attention is all you need,” Advances in neural information processing systems, 2017.