Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 4, Issue 1: 60-65, DOI: 10.25236/AJCIS.2021.040110

A Hybrid Sentiment Analysis Method of Transformer and Capsule Network for Hotel Reviews

Yanxiang Li, Jiajia Jiao

College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

Abstract: *As a significant branch of natural language processing, deep learning-based sentiment analysis has dominated in text sentiment analysis instead of lexicon graph and machine learning methods. To further improve the quality of sentiment analysis, we propose a hybrid sentiment analysis method of transformer and capsule network for hotel reviews. The proposed approach takes advantages of both self-attention mechanism in transformer and detailed representation in capsule network to capture bidirectional semantic features well. Compared with the traditional RNN, CNN and pure transformer, the hybrid sentiment analysis method of transformer and capsule network performs 13.83%, 8.97%, and 8.02% higher accuracy for an open-source dataset of hotel reviews respectively. The comprehensive experiments results demonstrate that our proposed method achieves higher quality of sentiment analysis than latest methods.*

Keywords: *Sentiment analysis, transformer, capsule network*

1. Introduction

Sentiment analysis is the computational study of people's opinions, sentiments, emotions, appraisals, and attitudes towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [1]. Sentiment analysis methods mainly include three categories: sentiment lexicon-based, machine learning-based and deep learning-based methods.

Sentiment lexicon-based method refers to a collection of words or phrases with emotional colors, these words can be adjectives, also can be nouns, adverbs and verbs, generally with one polarity, emotional words can generally be divided into positive and negative emotional words, emotional words are generally refers to a certain [2]. Statistical affective words can be extracted from text by algorithms or a pre-established affective dictionary can be used for searching and matching. However, this method relies heavily on the accuracy of sentiment word extraction or sentiment lexicon, and it takes a lot of time and energy to extract sentiment words or make sentiment lexicon [3].

Machine learning-based method is to manually extract text features and then the computer processes the text according to a specific algorithm and then outputs sentiment classification [4]. The key of machine learning is to select text features suitable for sentiment classification, and much subsequent work focuses on the selection of features [5]. Machine learning has significant advantages over approaches that rely entirely on manual constructs of sentiment lexicon. Support vector machine (SVM) and Bayesian algorithm are commonly used in machine learning methods. This kind of method generally constructs features artificially from the text with emotional labels, and selects the most effective features to represent the text through the method of expected cross entropy for training, and finally obtains the polarity of the text [6]. However, the construction of features relies more on a large number of manually extracted text features, and the whole process is greatly interfered by human factors [7].

In recent years, deep learning-based methods have attracted more and more attention. Deep learning can simulate the construction of human brain nervous system to gradually analyze texts, extract features, automatically learn and optimize texts, so as to improve the correctness of text classification [8] [9] [10]. Bahdanau [11] et al. adopted the attention mechanism, which is the first paper to propose the application of the attention mechanism to the field of NLP. In 2017, Vaswani [12] et al. proposed a sequence transformation model Transformer based on self-attention mechanism on this basis, which greatly improved feature extraction capability. Capsule network was originally a complex neural network structure used for image classification [13]. Capsule networks have been widely used in image processing because of their ability to learn the positional relationship. Lin Yue [14] et al. proposed a cross-domain sentiment classification method based on capsule network, which used capsule network for assistant

ISSN 2616-5775 Vol. 4, Issue 1: 60-65, DOI: 10.25236/AJCIS.2021.040110

classification. Xu Long [15] et al. proposed a short-text sentiment analysis method based on self-attention and capsule network, which combined self-attention and capsule network, thus greatly improving the accuracy of the model.

In this paper, we aim to tackle the problems above, and propose a hybrid sentiment analysis method of Transformer and Capsule network for hotel reviews. Transformer model can be used to better extract text feature information in the process of hotel review classification. Capsule network can make feature vector differentiation between different categories more obvious. The experiment shows that this method can effectively improve the performance compared with other methods in the comparative experiment.

2. Proposed hybrid model

In this paper, we aim to propose a hybrid sentiment analysis method of Transformer and Capsule network for hotel reviews. The method structure is shown in Fig.1.Firstly, preprocessing the dataset, such as word segmentation, to get the processed dataset. Then inputting the dataset into the Transformer encoder model to obtain the extracted feature information matrix T. Finally, feeding the matrix into the capsule network layer and getting the classification results.

2.1 Word vectors and location information

In this model, the input information is composed of word vector and position information. word vectors in the presence of large amounts of semantic information, so should sentiment analysis problem solving methods are word vector-based, then using the neural network model to mining text emotional tendency, this paper also follow this rule, but because of the Transformer encoder is a kind of completely the attention-based mechanism model. Therefore, the sequential relationship between words needs to be input into the coding network to learn the sequential feature information. The input value is the splicing value of word vector and position information, in which word vector is trained by Word2Vec model, skip-gram mode is adopted during training, window size is set as 5, vector dimension is set as 200, and position information is vectorized by one-hot coding.

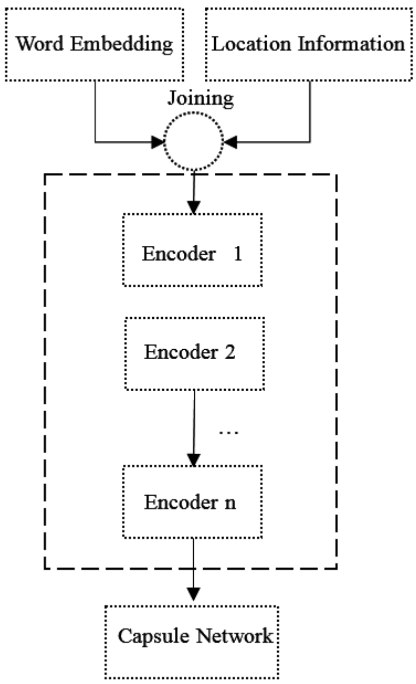

Figure 1: Model framework

2.2 Transformer Encoding

Transformer encoder consists of two sub-layers: the self-attention layer and the feedforward neural network layer. The residual network and normalization are added in each sub-layer. Among them, the multi-head attention model in the self-attention layer is composed of h Scaled Dot-Product Attention. The structure of multi-head attention is shown in Fig. 2.

ISSN 2616-5775 Vol. 4, Issue 1: 60-65, DOI: 10.25236/AJCIS.2021.040110

In Transformer, all words in the input sentence are processed at the same time without considering word ordering and position information, and the output is the same no matter how the sentence structure order is shuffled, so position encoding is introduced to solve this problem.

In the self-attention mechanism, each word has three vectors, namely Query vector (Q), Key vector (K) and Value vector (V)[16]. The multi-head attention mechanism projects Q, K and V through multiple different linear transformations, and finally splits together different results. The calculation formula is as follows:

$$M\_H(Q,K,V) = Concat(h1, h2, ..., hh)W^{o} \quad (1)$$

$$h_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (2)$$

$$Attention(Q,K,V) = soft\max(\frac{QK^T}{\sqrt{d_k}})V \quad (3)$$

As shown in Fig. 5, the calculation of single scaled dot product attention is relatively simple. If the Q, K and V are converted into matrices through linear calculation, then multiple scaled dot product units are used for calculation, and finally the calculated results are combined, more semantic features can be obtained.

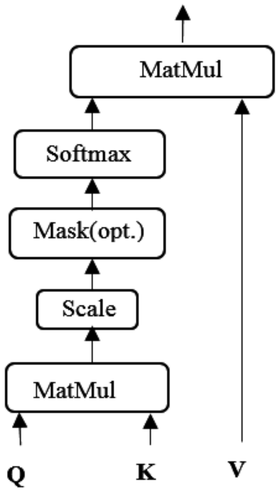

Figure 2: Scaled Dot-Product Attention

Firstly, the input corpus is converted into feature vectors by using the Word2Vect word embedding method in Transformer, and the location information of the binding words is incorporated into the input corpus. For text = {w_1, w_2... w_n}, the weight value is not 0 due to the introduction of position information. Therefore, its mask needs to be set to 0 before Softmax calculation to eliminate the influence of filling data. In this paper, the PADDING-MASK method is adopted to achieve the realization of the idea is to give a negative infinite value to the padding part, and then through the Softmax function to calculate the result is a value approaching 0.

The input matrix X is calculated according to Equations (1) and (2), and obtaining the matrix Z. Then the residual network and layer normalization are used to adjust the characteristic information.

Transformer encoders use multiple encoders to extract text features, and then input the learned features to the next level for emotional orientation classification.

2.3 Sentiment analysis

This layer is classified by capsule network, and its input is the output h_i of Transformer feature extraction layer, in which the model parameters are updated by dynamic routing algorithm [17] [18] [19], the module length of the capsule network is used to represent the probability of the corresponding category. Different capsule vectors represent different categories. The network structure diagram is shown in Fig.3.

The calculation formula of each parameter in the schematic diagram is as follows:

ISSN 2616-5775 Vol. 4, Issue 1: 60-65, DOI: 10.25236/AJCIS.2021.040110

$$u_{ji} = w_{ij} h_i \tag{4}$$

$$s_j = \sum_i u_{ji} \tag{5}$$

Normalize the final output, as shown in Formula (6):

$$v_i = \left(\frac{\|s_j\|^2}{1+\|s_j\|^2}\right) \times \left(\frac{s_j}{\|s_j\|}\right) \tag{6}$$

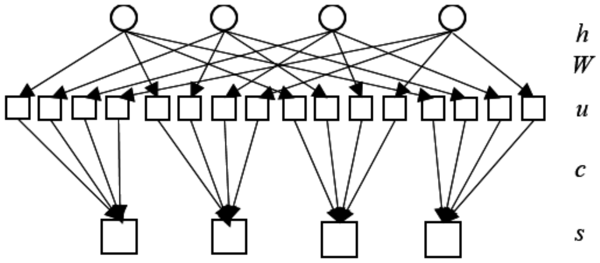

Figure 3: Capsule network layer structure

3. Experiments and Analysis

In this study, we use the corpus of hotel reviews organized by Professor Tan Songbo[1]. In order to explore the model performance on the small dataset, 6,000 balanced review data were selected, among which 3000 positive reviews and 3000 negative reviews were selected. Before being used as the input of the model, the data should be cleaned, the format and characters of the corpus data should be adjusted, the special symbols, space characters and newline characters unrelated to the data set should be removed, and the stop words should be processed. The training set is 70% and the test set is 30%, which is obtained by random segmentation of positive and negative samples.

The environment of this experiment is 64-bit Windows 10 operating system, PyCharm2017 and Jupyter Notebook, and the deep learning tool PyTorch is used to develop and train the model.

Experimental parameter Settings are shown in Table 1:

Table 1: Experimental Parameter Settings

Parameter	Value
Learning-rate	0.001
Batch_size	10
dropout	0.5
Num_head	5
Num_capsule	4
Share_weight	TRUE
Routing	3

3.1 Results analysis of large sample data sets

In the experiment, the evaluation metrics are Precision, Recall and F1 value. Accuracy rate is the correct prediction rate among all samples with positive prediction results, and recall rate is the positive prediction rate among samples with positive actual results. F1 value is the balance point between accuracy rate and recall rate [20] [21]·Since the index of F1 score is more balanced and representative, this paper adopts the value of F1 as the first evaluation index.

Table 2 shows the three main stream for deep learning-based model in this paper, the model of the results of the comparison, contrast can be found: under the condition of sample number (4800 training samples), four models are fairly good results have been achieved, the use of the proposed model of sentiment analysis based on network Transformer and capsule, the overall accuracy are improved

[1] https://github.com/zy1257

ISSN 2616-5775 Vol. 4, Issue 1: 60-65, DOI: 10.25236/AJCIS.2021.040110

obviously. It can be seen that the model proposed in this paper is effective and can effectively improve the accuracy and accuracy of text sentiment analysis.

Table 2: Text sentiment analysis results

Model	Precision	Recall	F1
RNN	0.8620	0.8467	0.8593
CNN	0.8863	0.8742	0.8802
Transformer	0.8864	0.8523	0.8602
Proposed model	**0.9028**	**0.9037**	**0.9034**

3.2 Results analysis of small sample data sets

In order to verify the effectiveness of the model proposed in this paper on small sample data, a small sample test was designed, namely training in 500 training samples and training in 200 training samples respectively. The experimental results are shown in Table 3 and Table 4.

Table 3: Sentiment analysis results of 500 training samples

Model	Precision	Recall	F1
RNN	0.7603	0.8019	0.7832
CNN	0.8000	0.8211	0.8100
Transformer	0.8204	0.8323	0.8280
Proposed model	**0.8523**	**0.8665**	**0.8598**

Table 4: Sentiment analysis results of 200 training samples

Model	Precision	Recall	F1
RNN	0.7023	0.7203	0.7104
CNN	0.7399	0.7524	0.7421
Transformer	0.7534	0.7423	0.7486
Proposed model	**0.8057**	**0.8154**	**0.8087**

As to small sample datasets, the accuracy of mainstream deep learning models decreases by 10%-20%, but the improvement is larger comparing with the traditional model. On the contrary, traditional deep learning models does not perform well, mainly because traditional models cannot be effectively trained on small datasets and cannot effectively capture semantic features of text. However, the method proposed in this paper is stable on small sample datasets, and performs better than the traditional deep learning-based model on both 500 and 200 datasets. Therefore, Transformer layer fully extracts semantic knowledge from samples, while the capsule layer can classify them well. Therefore, the method proposed in this paper still has good performance on small sample datasets.

4. Conclusion

In this study, for a higher accuracy, a hybrid sentiment analysis method of Transformer and Capsule network for hotel reviews is proposed. This model not only makes full use of the complete information, reducing the distance between the interdependent features, but also can extract more abundant text information. Under the same training conditions, Compared with the traditional RNN, CNN and pure transformer, the hybrid sentiment analysis method of transformer and capsule network performs13.83%, 8.97%, and 8.02% higher accuracy for an open-source dataset of hotel reviews respectively. At the same time, a comparative experiment on a small sample data set shows that the model has an excellent performance on a small sample data set and improved the accuracy.

References

[1] Zhang L, Wang S, Liu B. Deep Learning for Sentiment Analysis: A Survey [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(4): 1253-1287.
[2] MEI L L, HUANG H Y, ZHOU X Y, MAO X L. An overview of sentiment lexicon construction [J]. Journal of Chinese Information Processing, 2016, 30(5): 19-27.
[3] Schuller B, Mousa E D, Vryniotis V. Sentiment analysis and opinion mining [J]. Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery, 2015, 5(5): 255-263.

ISSN 2616-5775 Vol. 4, Issue 1: 60-65, DOI: 10.25236/AJCIS.2021.040110

[4] HONG W, LI M. A review of textual sentiment analysis methods [J].Computer Engineering and Science, 2019, 41(04): 750-757.
[5] DU H, XU X K, WU D Y, LIU Y, YU Z H, CHENG X Q. A Sentiment Classification Method Based on Sentiment-Specific Word Embedding [J]. Journal of Chinese Information Processing, 2017, 31(3): 170-176.
[6] MIAO G H. Emotion Mining and Simulation Analysis of Microblogging Based on Word2vec and SVM [J]. Electronic Science and Technology, 2018, 31(5): 81-83.
[7] CEHNG Z S, WANG L. Sentiment analysis method of network comments based on support vector machine [J]. Electronic Technology & Software Engineering. 2019, 162(16): 19-20.
[8] SHEN Y L, ZHAO X B. A Review of Research on Aspect-based Sentiment Analysis based on Deep Learning [J]. Information Technology & Standardization, 2020, (Z1): 50-53+58.
[9] PING Y N. Fine-grained Sentiment Analysis of Multi-source News Comments Based on Capsule Network [D].Shanghai: Shanghai Normal University, 2020.
[10] Kim Y. Convolutional Neural Networks for Sentence Classification [J]. Eprint Arxiv, 2014.
[11] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate [J]. Computer Science, 2014.
[12] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need [J]. arXiv, 2017.
[13] WU Y W, HANG K, WANG X Y, et al. Method of Emotional Classification in Short Texts Combined with LDA Models [J]. Journal of Chinese Computer Systems, 2019, 40(10). 2082-2086.
[14] WANG J Q, GONG Z H, XUE Y, PANG S G, GU D H. Target-specific Sentiment Analysis Based on Mixed Multi-head Attention and Capsule Network [J]. Journal of Chinese Information Processing, 2020, 34(5): 100-110.
[15] SABOURS, FROSSTN, HINTONGE. Dynamic routing between capsules [C]. Advances in Neural Information Processing Systems. 2017: 3856-3866.
[16] LIN Y, QIAN T Y. Cross-domain Sentiment Classification Based on Capsule Network [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2019, 11(03): 286-294.
[17] XU L. Short Text Sentiment Analysis Based on Self-attention and Capsule Network [J]. Computer and Modernization, 2020 (07): 61-64+70.
[18] ZHONG M H, HUANG H. An analysis of the connotation of socialist core values [J]. Shandong Social Sciences, 2009 (12): 14-18.
[19] HU Y J, JIANG J X, CHANG H Y. A New Method of Keywords Extraction for Chinese Short – text Classification [J]. Data Analysis and Knowledge Discovery, 2013 (6): 42-48.
[20] HINTONGE, SABOURS, FROSSTN. Matrix capsules with EM routing [C]. The 6th International Conference on Learning Representations (ICLR). 2018: 789.
[21] LIU L F, YANG L, ZHANG S W, LIN H F. Microblog Sentiment Analysis Based on Convolutional Neural Network [J]. Journal of Chinese Information Processing, 2015, 29(6): 159-165.