WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Research Progress of Sentiment Analysis Based on Deep Learning

Jiameng Li *

Intelligent science and technology, Communication University of China, Beijing, China

* Corresponding Author Email: jiamengli@cuc.edu

Abstract. Sentiment analysis technology is widely used in film and television media, social media, entertainment industry, medical and other fields, and has great commercial use value. In the early stages of research, sentiment analysis methods based on traditional machine learning are not effective, which greatly limits its application in the commercial field. In recent years, with the development of artificial intelligence, deep learning-based sentiment analysis have made great advances. However, the large variation in performance results and characteristics for different algorithms heightened the difficulty in practical application. According to the technical progress of sentiment analysis, this paper summarizes from two aspects: data set and algorithm, introducing the commonly used data sets, evaluation indicators and sentiment analysis methods based on deep learning. In addition, this paper also provided a comparison among different algorithms widely used for sentiment analysis. Finally, this paper analyzes the existing technical adjustment, and discusses the future research direction.

Keywords: Sentiment Analysis, Deep learning, Natural language processing, LSTM, Transformer.

1. Introduction

Sentiment Analysis, also known as opinion mining or emotion AI [1], is an important research field of computational linguistics and natural language processing (NLP), which aims to identify and extract subjective information and mine underlying thoughts and emotions from the text [2]. It categorizes a given text into one or more emotional categories and analyze its tendencies, such as positive, negative, or neutral. Advanced sentiment analysis techniques can also identify and further classify emotions and opinions (e.g., happiness, sadness, fear, anger, etc.).

Sentiment Analysis has a wide range of applications including Internet interaction, marketing, health care, and politics.

Use sentiment analysis to monitor mainstream opinion on social media, support client relations supporting system and manage customer feedback.

Use sentiment analysis to speculate user preferences and push content that are more interesting and products that meet customer needs to users.

Use sentiment analysis to analyze the impact of specific products or brands on consumers, understand consumers' evaluation and needs in order to addressing product limitations.

Use sentiment analysis to judge the physical health of patients in the course of treatment.

Use sentiment analysis to analyze real-time reactions on social media brought about by political campaigns and creative policies.

All of the applications above help researchers, enterprises and governments to find out the state of public opinion more accurately, assisting them to make more informed decisions and adjusting operational methods and direction promptly.

However, as a branch in the field of NLP, sentiment analysis still has challenges ahead and the following are the key pain points.

Ambiguity and contextual dependence: The meaning of words and phrases is highly dependent on context, making the accuracy of sentiment analysis more difficult to guarantee when encountering ironic, figurative, and ironic texts.

Language nuances and domain particularity: Sentiment analysis techniques need to be adapted to specific fields and industries to deal with specialist vocabulary and stereotypical industry expressions, and language nuances such as slang and regional dialects also pose significant challenges for sentiment analysis techniques.

Limited labeled data: Supervised learning techniques rely on large labeled data sets, which has to take into account the high time and cost of creating these data sets. This is especially challenging for low-resource languages or specialized fields [3].

Multilingual sentiment analysis: With the continuous development of the Internet and the deepening of cultural exchanges between countries, language texts have become more diversified, and the development of models dealing with multiple languages will also be a continuous challenge in the area of sentiment analysis.

In this paper, the relevant data and algorithms of sentiment analysis are first outlined. The second chapter recommends relative evaluation index and three datasets, including ChnSentiCorp dataset, MPQA dataset and Amazon dataset. The third chapter presents detailed descriptions of the deep learning models (CNN, RNN, LSTM) and algorithms (LSTM, Transformer). Finally, the fourth chapter is the sum and prospect.

2. Metrics and Datasets for Sentiment Analysis

2.1. Metrics

For the sentiment analysis model, Accuracy, Recall, Precision and F1 value (F1_Score) are often selected as the evaluation indexes of the model [4]. The following uses the confusion matrix as shown in Table 1 to briefly introduce the above indicators, as follows:

Table 1. Label and prediction comparing

Label \ Prediction	Positive(P)	Negative(N)
Ture(T)	TP	FN
False(F)	FP	TN

Accuracy. Measure the proportion of the total sample with correct predictions [5]. Although the accuracy rate can judge the comprehensive accuracy rate from an overall perspective, if the balance of the sample cannot be guaranteed, the calculation of a high accuracy rate is meaningless and cannot be used as a good indicator to weigh the result and assist the judgment, and the accuracy rate is invalid.

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

Precision. Measure the proportion of labels correctly assigned by the system [6]. The ratio of the predicted true positive samples to all predicted true samples.

$$\mathrm{precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

The precision represents the accuracy of the prediction in the positive sample results, and the accuracy represents the accuracy of the overall prediction, including positive and negative samples.

Recall. Measure the proportion of tags found by the system. Based on actual samples, the ratio of the number of samples that are correctly predicted as positive examples (True Positives, TP) to the number of samples that are actually positive examples (True Positives + False Negatives, TP + FN). The recall rate indicates how many truly positive cases have been successfully predicted.

$$recall = \frac{TP}{TP + FN} \tag{3}$$

F1 score. The F1 score is the harmonic average of accuracy and recall and is used to consider the performance of the classifier. It is calculated as follows:

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall} = \frac{2TP}{2TP + FP + FN} \tag{4}$$

F1 scores range from 0 to 1, with 1 representing the perfect classifier and 0 representing the worst. The F1 score is particularly useful in the case of dealing with unbalanced categories because it takes into account false positive and false negative cases, not just accuracy. In some cases, there is a trade-off between accuracy and recall, and the F1 score helps to find a balance that allows the classifier to achieve good performance between accuracy and recall [7].

2.2. Datasets

Sentiment analysis has a widespread adoption in hotel reviews, sales websites, news articles and other fields, among which the commonly used data include ChnSentiCorp data set, MPQA data set, Amazon data set and so on.

ChnSentiCorp data set, published by Tan Songbo, is a commonly used data set for public Chinese sentiment analysis. Collected from a large-scale hotel review, it contains more than 7,000 hotel review data, including positive and negative emotion types.

The MPQA data set, a multi-perspective question-and-answer data set, is a two-category labeled opinion corpus collected from news articles related to a variety of news sources and contains 10,606 pieces of data. It is an unbalanced data set consisting of 3311 positive samples and 7293 negative samples.

The Amazon data set is a popular product review corpus data set, collected from the Amazon website, which contains binary classification and five-classification labels, the binary classification data set contains 36 million training samples and 400,000 test data [8]. The five-category version contains 3 million training samples and 650,000 test samples.

3. Deep Learning Models for Sentiment Analysis

In the process of emotion analysis, deep learning models cannot be avoided. The widely used deep learning models include recurrent neural networks (RNN), convolutional neural networks (CNN), long short-term memory models (LSTM).

3.1. Basic Network Structure

3.1.1. CNN

The typical structure of CNN includes convolution layer, pooling layer, fully joined layer and so on [9]. It is a deep learning model commonly used in image and video processing. Compared with the traditional neural network, CNN makes a major breakthrough in processing image and sequence data, it can automatically learn the features in the image and extract the most useful information. The core feature of CNN is convolution operation, which can calculate the sliding window on the image and extract the features of the image through filters and pooling layers, as shown in Fig 1. Therefore, CNN has a variety of applications in image classification, target detection, speech recognition and other fields.

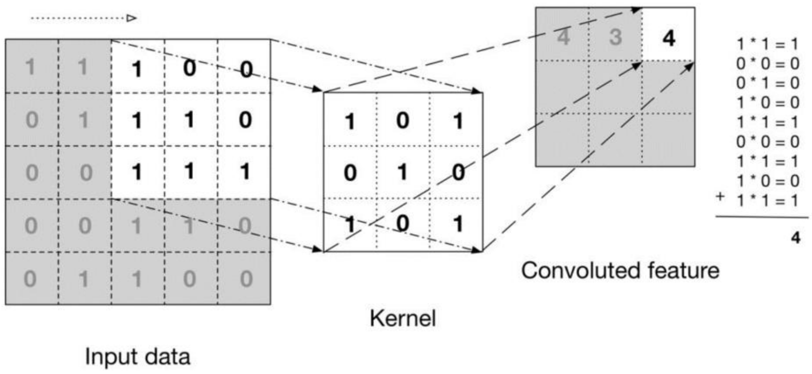

Fig. 1 CNN structure [10]

3.1.2. RNN

RNN is composed of an input layer, a hidden layer and an output layer, and is a kind of neural network, as shown in Fig 2. RNN is widely used to mine the timing information and semantic information in the text, especially highly effective for data with sequence characteristics [11]. The deep learning model has made breakthroughs in solving NLP problems such as speech translation, keyword spotting, language construction, neural machine translation and time-series analysis.

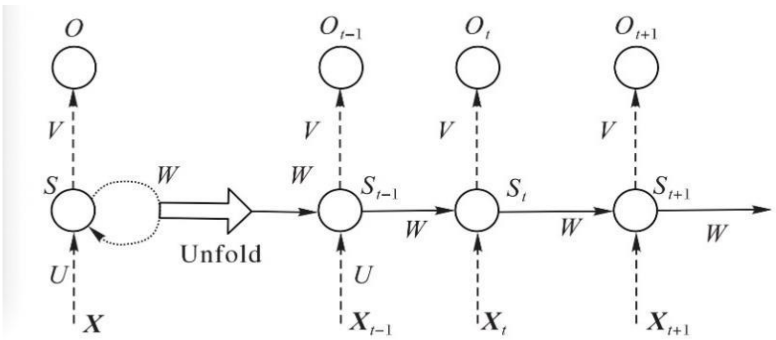

Fig. 2 RNN structure [10]

3.1.3. LSTM

LSTM adds memory unit c, input gate I, forgetting gate f, and output gate o in comparison to simple recurrent neural networks. Recurrent neural networks are much better able to process lengthy data sequences when these gates and memory units are combined. As an enhanced recurrent neural network, it is highly efficient in processing sequence data and can not only address the issue that RNNs cannot manage long distance dependence, but it can also address the issue of gradient explosion or disappearance, which is a typical problem in neural networks. LSTM has the potential to produce the most accurate results, while preserving the context of the text or corpus, and greatly reducing the effort of feature extraction when the neural network formed by learning important features. However, LSTM inevitably requires large scale valid training data and feature extraction, high training and operation costs, and low interpretability. In addition, it is the most expensive method in terms of computational time, especially if word embeddings are trained jointly, as it requires a large dataset.

3.2. ·Sentiment Analysis based on CNN

In the early stage of sentiment analysis research, the word vector representation method based on machine learning is less effective in sentiment analysis. He et al. proposed a MCNN sentiment analysis method based on word vector and deep learning model [12]. As shown in Figure 1, the algorithm is divided into three modules. First, the word vector model is used for word segmentation and then vector representation of the model is carried out. Then the emotion matrix is constructed to realize the mapping between word meaning and emotion space. Finally, the neural network model is used to complete the classification of the vector to the emotion category. While ensuring the classification accuracy, it greatly improves the speed of emotion classification. Although this method achieves good results, the effect of the model is greatly affected by the word vector model. In addition, the neural network model cannot optimize the network parameters together with the word vector model, which greatly limits the effect of the model.

3.3. ·Sentiment Analysis based on CNN-LSTM

With the continuous development of the deep learning field, Jin et al. proposed a dimensional sentiment analysis method based on the local CNN-LSTM model [13]. The model is composed of local CNN and LSTM, which can capture the remote dependencies between sentences while capturing the local information inside the sentences. The algorithm is divided into four parts: First, word embedding is used to construct word vectors for words in the vocabulary; Then, the local CNN is used to construct text vectors for the specified text -- the text is divided into individual sentences into regions in order to extract useful affective features in different regions and evaluate their contribution to VA prediction; Then, the word vectors go through the convolution layer and the maximum pooling layer successively. Finally, LSTM integrates these local regional features in turn across regions to build text vectors for VA prediction. Dimensional sentiment analysis can be implemented by identifying the VA ratings of texts and ranking them to prioritize highly evoked negative emotions in the system, thus providing smarter and more accurate new granular sentiment applications. The laboratory results demonstrate the significant superiority of this strategy over the conventional dictionary-based, regression-based, and neural network-based approaches suggested in earlier research.

3.4. ·Sentiment Analysis based on Transformer

In previous sentiment analysis studies of tweets, Word2Vec and GloVe were unable to take into account the polysemy and noise contained in tweets. Considering language ambiguity and ambiguity, specific context, word emotion, grammatical knowledge and data cleanliness, Usman Naseem et al. proposed a deep intelligent Context Embedding (DICE) sentiment analysis framework for Twitter based on transformer architecture [14]. An intelligent preprocessor, a text presentation layer, and a bidirectional long short-term memory with attention modeling (BiLSTM) are the core components of the framework. To eliminate noise, the intelligent tweet preprocessor performs spelling correction, emotional perception tokenization, word segmentation, and normalization before implementing deep context embedding. The quality of the tweet is improved by dealing with noise in the context, and finally, the improved word representation is forwarded to BiLSTM, which uses the BiLSTM to determine the mood of the tweet. The algorithm's proposed Transformer-based word representation technology incorporates deep intelligent context embedding, which can capture ambiguity in context, represent diverse properties of words (including their syntax, semantics, and emotion), and effectively deal with complex properties of words, their use in the highly complicated environment of tweets, and other relationships in deeper level. DICE can also deal with linguistic ambiguity by learning word representations, including ambiguity, semantics, grammar, OOV words, and emotional knowledge. Through the use of airline's tweets, practice shows that the framework has a distinct advantage with the current SOTA in terms of sentiment classification.

4. Conclusion

This paper reviews the development of deep learning models in the field of sentiment analysis, investigates the current development status of sentiment analysis, and then outlines the commonly used data sets and evaluation indicators for sentiment analysis. Finally, it introduces the main deep learning methods and the innovation points, advantages and challenges of the current three advanced sentiment analysis algorithms.

In the future, as deep learning models continue to evolve, sentiment analysis will need to further reduce the need for large data sets and reliance on manual training data. In addition, sentiment analysis should carry out more targeted fine-tuning in a smaller field, that is, the fine-tuning model of a large model in a specific scenario will better solve the special problems of the field, improve efficiency, and achieve the balance between optimization accuracy and storage efficiency, which should be the focus of research.

References

[1] Jinpei Yan, Yong Qi, Qifan Rao.Detecting Malware with an Ensemble Method Based on Deep Neural Network, IEEE International Conference On Big Data Science And Enginerring,2018, 1170-1179.

[2] Rehab Duwairi and Mahmoud El-Orfali. A Study of the Effects of Preprocessing Strategies on Sentiment Analysis for Arabic Text, Journal of Information Science, 2014, 40(4):501-513

[3] Cházaro-Ruiz LF, Kellenberger A, Dunsch L. Dunsch. In situ ESR/UV-vis-NIR and ATR-FTIR spectroelectrochemical studies on the p-doping of copolymers of 3-methylthiophene and 3-hexylthiophene, Journal of Physical Chemistry, 2009, 113(8): 2310-2316.

[4] Mathias Kraus, Stefan Feuerriegel. Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees, Expert Systems with Applications,2019, 65-79.

[5] Joeky T Senders, Omar Arnaout, Aditya V Karhade, et. al. Natural and Artificial Intelligence in Neurosurgery: A Systematic Review, Neurosurgery, 2018, 83(2):181-192.

[6] Author Picture Rathany Chan Sam, Author Picture Huong Thanh Le, Author Picture Thuy Thanh Nguyen, et. al. Combining proper name-coreference with conditional random fields for semi-supervised named entity recognition in Vietnamese text, Proceedings of the 15th Pacific-Asia conference on Advances in knowledge discovery and data mining - Volume Part I, 2011, 512–524.

[7] Resch, Bernd, Usländer, Florian, Havas, Clemens. Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment, Cartography and Geographic Information Science, 2018, 362-376.

[8] Vineet John, Olga Vechtomova. Sentiment Analysis on Financial News Headlines using Training Dataset Augmentation. Proceedings of the 11th International Workshop on Semantic Evaluation, 2017.

[9] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan. CNN Features off-the-shelf: an Astounding Baseline for Recognition, 2014 IEEE conference on computer vision and pattern recognition workshops,2014, 512-519.

[10] WANG Yingjie, ZHU Jiuqi, WANG Zumin1, BAI Fengbo, GONG Jian. Review of applications of natural language processing in text sentiment analysis, Frontiers in Business Economics and Management, 2023, 10(1):58-64.

[11] Adji B. Dieng, Chong Wang, Jianfeng Gao, John Paisley. TopicRNN: a recurrent neural network with long-range semantic dependency, Cornell University, 2016.

[12] He Yanxiang, Sun Songtao, Niu Feifei. A Deep learning model of emotion semantic Enhancement for Weibo sentiment Analysis. Journal of computers, 2017, 40 (4): 18.

[13] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao, Attention-based LSTM for Aspect-level Sentiment Classification, Association for Computational Linguistics,2016, 606–615.

[14] Usman Naseem, Imran Razzak, Katarzyna Musial, Muhammad Imran. Transformer based Deep Intelligent Contextual Embedding for Twitter sentiment analysis. Future Generation Computer Systems, 2020, 58-69.