Deep Learning Techniques for Sentiment Analysis

Erjia Fu[1,a,†], Junyuan Xiang[2,b,†], Chuanhao Xiong[3,c,*,†]

[1] Wuhan University, China

[2] UW-Madison

[3] Beijing Normal University, China

[a]544257169@qq.com, [b]jxiang34@wisc.edu, [c]1901090006@mail.bnuz.edu

* Corresponding Author Email: 1901090006@mail.bnuz.edu

† These authors contribute equally

Abstract: Sentiment analysis covers a wide range of computational research, including research on the opinions, feelings, emotions, evaluations of people, and attitudes toward products, services, organizations, individuals, issues, events, topics, and their attributes. It plays an increasingly important role in the era of big data. In fact, it has spread from computer science to management and social sciences such as marketing, finance, political science, communications, medical science and even history, generating common interest throughout society due to its commercial importance. TEA is a basic task with the typical used of NLP methods which is full of interest, particularly for fine-grained classification of textual emotional content. It is the process of mastering, inductive analysis and reasoning about emotional content. Simply put, it is the process of analysing, processing, summarising and reasoning about emotive and subjective texts. The Internet generates a large numberof user reviews to gain valuable information about people, events and products. These reviews express a wide range of emotions and emotional tendencies, including joy, anger, sadness, delight, criticism and praise. Potential users can therefore view these subjective reviews to understand how public opinion views an event or product.

Keywords: Deep Learning, Sentiment Analysis, Big Data, Neural Network.

1. Introduction

Sentiment Analysis is a popular topic of research at domestic and international level in recent years. It undertakes the task of helping users quickly obtain, organize and make an analysis about the relevant assessment information, and to analyze, process, generalize and draw inferences from subjective texts with emotional overtones.

Multiple tasks, for instance, sentiment classification, opinion extraction, question answering and sumarization together constitute sentiment analysis [1]. As a result, it is of great difficulty to merely classify it as a certain field, and it is often classified in different directions from different angles. If you simply discriminate the tendency of text, it can be regarded as a classification task; if you want to extract relevant elements (opinion holders, opinion evaluation objects, etc.) from opinion sentences, it is an information extraction task; and if to find an opinion on something from a huge amount of text can be viewed as a retrieval task.

As Internet technology is growing in speed and popularity, people are increasingly concerned about web information’s subjective tendency classification with the increasing demand for web page management, and for supervision and blocking of harmful (or spam) information. There are some differences between this classification and the traditional one. Traditional text categorization focuses on the objective content of the written text [2], while the object studied by the tendency classification is the "subjective factor" of the text, that is, the subjective tendency expressed by the author. The result is for a particular text to get information on whether it supports a certain point of view. This unique text classification task is also known as sentiment classification.

Affective categorization means classifying a text into a variety of praise and criticism based on the meaning and emotional information conveyed by the text, and is a segmentation of the inclination, opinion and attitude of the author, so it is sometimes referred to as opinion analysis.

As a particular taxonomy of issues., sentiment classification has both Frequently asked questions about the normal model classification and its particularities, For example, concealment, ambiguity and Polarity of affective message performance.

A lot of research has been done on these problems, and many classification methods have been proposed. These methods can be classified by machine learning methods or by the characteristics of emotional texts.

According to the labeling of training samples used by machine learning methods, sentiment text classification can be roughly divided into three categories: supervised learning methods, semi-supervised learning methods and unsupervised learning methods [3].

Supervised learning methods: Sentiment classification methods based on supervised learning use machine learning methods for training a large number of labeled samples. In 2002, supervised learning methods were first employed for sentiment classification. The impact of various classification algorithms and a variety of characteristics as well as feature weights selection methods in sentiment classification on the basis of supervised learning were compared in the literature. In 2004, subjective sentence summaries were introduced into sentiment classification; in 2010, the impact of polarity transfer on sentiment classification was analyzed; in 2011, the performance of sentiment classification was effectively improved by using an ensemble learning method based on feature space and classification algorithms.

Semi-supervised learning methods: Sentiment classification methods based on semi-supervised learning build classification models by training on some labeled samples and learning on many unlabeled samples. Multiple machine learning methods (for instance clustering methods, ensemble learning, and so on.) were integrated into sentiment classification on the basis of semi-supervised learning in 2009; in the face of the lack of Chinese labeled corpus in sentiment classification, in 2009, a collaborative learning method was used to use labeled English corpus [4]. and unlabeled Chinese corpus to achieve high-performance Chinese sentiment classification. 2010 divided the expression of sentiment text into two views, personal and impersonal, and applied collaborative learning for semi-supervised learning for sentiment classification.

Unsupervised learning methods: Sentiment classification methods based on unsupervised learning refer to only using unlabeled samples for sentiment classification modeling. Most of the previous research work is to achieve unsupervised classification through the seed word set marked by sentiment classification. In 2002, the sentiment tendency of the text was calculated by calculating the point mutual information between the candidate words and the seed sentiment words in the text." and "poor" are used as seed words. After obtaining the point mutual information between each word and the seed word, the sentiment tendency of each word is calculated according to SO-PMI, and the overall sentiment tendency of the text is calculated by word count. sex. 2006 Extract words' sentiment information via HowNet-based semantic analysis. 2009 In the sample space, a shallow semantic analysis method based on "Latent Dirichlet allocation, LDA" obtains the labels of unlabeled samples based on the co-occurrence relationship between documents and words.

2. Background

2.1. Task Definition

Scores of people have been observed to engaged in social media zealously and actively. People express their opinions in comments, posts and other forms on various topics. Consequently, the Internet produces a wealth of data that can be used for in-depth research. This makes sentiment analysis a hot field with a wide range of applications. For example, a movie's quality can be seen through sentiment analysis of a large number of audience reviews. Hence, to demonstrate the motivation behind the sentiment analysis ,we briefly show some of the works in the field of sentiment analysis.

2.2. Datasets

Data sets play an important role in testing the performance of sentiment analysis algorithms. We have identified a number of popular sentiment analysis datasets

ALI TOOSI: In tweets, the dataset contain hate speech. For simplicity, if a tweet is racist or sexist, we say it indludes hate speech. As a result, the task is to classify these tweets from other tweets.

Twitter US Airline Dataset: This data set contains tweets from all major U.S. airlines since February 2015. It includes Twitter user IDS, emotional confidence scores, positive and negative causes, retweet counts, tweets, dates, times and locations.

Financial Sentiment Analysis: The following data are intended to advance financial sentiment analysis research. It is combines two datasets into one easy-to-use CSV file. It provides economic sentences with emotional labels.

Therefore, the sentiment analysis data set mainly includes: tweets, debates, messages, comments, blogs, posts, consumer reviews, emojis and images, good accounts and bad debts. The highest accuracy obtained on IMDB dataset using CNN + RNN is 93.2%, while the highest accuracy obtained on Amazon multi-domain dataset using CNN + BI-LSTM + Attention model is about 87%. Therefore, RNN and its variants (such as LSTM, etc.) provide high accuracy for the task to analyze sentimental information, which demonstrates the popularity of RNN in data. In addition, the combination of CNN + RNN is used for visual sentiment analysis of image and video related data.

2.3. Metrics

A great majority of the famous sentiment analysis uses accuracy, F1 score, precision, and recall as parameters to measure performance. These parameters can be defined as below:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F1score} = \frac{2*Precision*Recall}{Precison+Recall}$$

Where TP is the true positive, FP is the false positive, TN is the True Negative, and FN is the False Negative.

In addition, other factors, for instance ranking loss, etc., and Rand Index were used to interpret the results correctly

$$\text{MeanSquareError(MSE)} = \frac{\sum_{i=1}^{N}\left(Actual_i - Predicted_i\right)^2}{N}$$

To gauge the difference between predicted and actual sentiment labels we use MSE (Verma et al., 2018). we minimize the MSE to train Sentiment models (Jiang et al., 2014; Wang et al., 2018c)

$$\text{Rankingloss} = \sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{m*n}$$

To measure the average distance between the true and predicted sentiment values for the m sentiment categories in the n test samples, we use Ranking loss (Moghaddam and Ester 2010)

$$\text{RandIndex}(C_i, C_m) = \frac{2(x+y)}{k*(k-1)}$$

Where x and y denotes the number of pairs assigned to the same and different clusters, and y denotes the number of respectively. $C_i\, and\, C_m$ denotes the clusters formed by model I and by manual annotation m, k denotes count of detected aspects.

RandIndex is commonly used in clustering problems to determine the similarity between two data clusters. The value is between 0 (no common pair of points for data clustering) and 1 (data clustering is identical).

3. Recent Researches & Neural Network

This thesis puts forward an improved method to represent words, that incorporates the contribution of emotional information into the traditional TF-IDF algorithm and generates weighted word vectors which are input into BiLSTM to capture contextual information in an effective manner [5]. A feedforward neural network classifier is used to obtain the comment's sentiment tendency. The sentiment analysis method put forward in this thesis is compared with other traditional methods such as RNN, etc. under the same conditions. It can be concluded that the method to analyze sentimental information has higher precision, recall, and F1 score according to the experiment. However, the degree of word discrimination will be weakened accordingly if the word vector is input into the BiLSTM model to analyze sentimental information, the accuracy of this analysis will be reduced. To some extent, the method to represent word put forward in this thesis alleviates the above problems by considering the sentimental information included in the words and its contribution to the task. We can gain the sentimental information included in the text through softmax mapping as well as the feedforward neural network. The effectiveness of the word representation method put forward in this thesis can be confirmed by the experiments with different word representation methods. It can improve the accuracy of the comment sentiment analysis method put forward in this thesis through the comparison experiments with other traditional methods. Nevertheless, in the training model, this method based on BiLSTM takes a long time.

In the field of natural language processing and data mining, sentiment analysis has always been a research hotspot. Recently, Deep Neural Network (DNN) models, LSTM models. What's more, people are paying more and more attention to their variants for instance GRU. Although they can be processed with these models regardless of the length of the sequence, they suffer from two drawbacks. 1. The feature space will become a high-dimensional space if they are used in the DNN's feature extraction layer. This model considers different features to be of equal importance [6].

To deal with these issues, the authors put forward ABCDM. This model extracts past and future context by taking the information flow in both directions into account with the utilizing of two independent bidirectional LSTM as well as GRU layers. Furthermore, applying the attention mechanism to the output of the bidirectional layer of ABCDM to emphasize different words more or less. By utilizing convolution as well as pooling mechanisms, this model can cut down features' dimensionality and extract location-invariant local features. In sentiment analysis, evaluating ABCDM's effectiveness is the most common and fundamental task. Experiments are carried out on 5 reviews as well as 3 Twitter datasets. This model achieves advanced results in both long comment and short tweet polarity classification according to the results by comparing this model with 6 recently proposed DNNs to analyze the sentimental information.

SKEP [7], named Sentiment Knowledge Enhancement Pre-training Model, is a fusion of multiple sentiment knowledge, including different domains and different types of knowledge, so that the same pre-training model can be used to handle multiple tasks of sentiment analysis. The starting point of the paper is that the current sentiment task must be divided into different domains using different models is cumbersome, as well as the discovery of aspects - sentiment words as a token when the effect will be better than splitting separately as 2 tokens, thus the three objectives optimization into the transformer, to achieve the model performance improvement, in most data sets are more than the previous SOTA.

Existing work in aspect-based sentiment analysis (ABSA) employs a unified approach that allows for interactive relationships between sub-tasks. However, we observe that these approaches tend to predict polarity based on the literal meaning of aspect and opinion terms and primarily consider implicit relationships between sub-tasks at the word level. Moreover, identifying multiple aspect-opinion pairs with polarity is more challenging. Therefore, a comprehensive understanding of contextual information about ABSA further requires aspects and opinions. In this paper, we propose the Deep Contextual Relationship Awareness Network (DCRAN), which allows interaction between sub-tasks with deep contextual information based on two modules (i.e., aspect and opinion propagation and explicit self-supervision strategies). In particular, we design novel self-supervised strategies for ABSA that have advantages in handling multiple aspects. Experimental results show that DCRAN significantly outperforms pre previous state-of-the-art approaches in three widely used benchmark tests. However, the current problem is that the extant methods only consider word-level relationships between subtasks and do not explicitly exploit the contextual information of the whole sequence

Analyze aspect-level sentimental information through convolution over dependency tree

The problems in this paper are based on the neural network (previous) approach ignoring information resources such as dependency trees that can shorten the distance between the aspect and opinion words, thus enabling effective retention of dependency information in lengthy sentences. In this paper, we propose the convolution of CDT model that uses Bi-LSTM to learn the representation of sentence features and further enhances the embedding (GCN) by graph convolutional networks to operate on the dependency tree of sentences in a direct manner. The model structure here is similar to the BilSTM model structure, except that the Attention mechanism is removed and the average pooling is added, and the experimental effect is improved. Experiments are conducted on four datasets, Rest14, Laptop14, Rest16, and Twitter, and the effects are compared with other benchmark models. The experiments prove that our method is the best method for aspect-b ased sentiment classification. Summary and reflection In this paper, we found that the architecture of CDT allows BiLSTM to interpret contextual information between consecutive words, while GCN enhances the embedding by modeling dependencies along the syntactic path of the dependency tree, and then validated against this idea, finally proving the feasibility of the method.

In the paper, in order to identify opinion words' sentiment polarity expressed in specific aspect of sentences, a method on the basis of neural networks is proposed. They present a convolution over CDT model that exploits a Bi-LSTM to learn the ways to represent the characteristics of a sentence, and GCN that operates on the dependency tree of the sentence in a direct manner is adopted to further enhance the embeddings. Their methods provide discriminative properties for supervising by propagating both context and dependency information from opinion words to aspect words. The integration with a GCN and a simple BiLSTM model, with the aim to capture structural and contextual information of sentences, shown that the GCN successfully performs convolutions on the dependency tree to refine BiLSTM embeddings. the model they propose is simple and out performs more complex and recent models tackling the same problem.

There are currently two deep learning algorithm models for word embedding [8]: Word2Vec and Global Vectors (GloVe). Although these two models are quite successful, some limitations need to be improved: (1) The size of the corpus needed has to be large because of the need to train and propose an acceptable vector for all the words. For example, to train the Word2Vec algorithm, Google has used nearly 100 billion words and has republished a pre-trained word vector with 300 dimensions. When the size of the dataset under study is small, researchers are forced to use these vectors, for instance Word2Vec, but the most suitable data for them to use for their studies might not be obtained in this way. (2) The word vector computations for both methods used to represent documents do not take into account the document's context. For instance, the "beetle" vector as a car is equal to the animal vector. (3) In the case where multiple terms do not appear at the same time, the two models ignore the relationship between them. (4) When sentiment analysis is performed on word vectors with sentiment polarity, it will result in misjudgment by ignoring the sentimental information of a specific text.

4. Challenge

Emotional AI developers still have several challenges to overcome [9].

A common concept in machine learning these days is that the success of emotional AI "training" always depends on the quality of the input data. Larger, better, and cleaner datasets are necessary to avoid the "garbage in, garbage out" situation, which poses challenges such as:

Challenges in text sentiment analysis: inability to identify double meanings, jokes, and innuendo; inability to account for regional differences in linguistic and non-native phonetic structures.

Handling all kinds of rhetoric in written material is a big challenge for Sentiment AI. They have difficulty understanding sarcasm or mood particles with different meanings, which means that Sentiment AI's analysis can be very different from what the original text is trying to convey. Sentiment AI cultivated in such an environment is extreme and inaccurate.

Example: "AI chatbot turns racist after spending a day on Twitter"

Challenges of Visual Sentiment Analysis: Inability to distinguish between real and forced or exaggerated emotional expressions; does not include body language; issues with processing conceptual and abstract images.

Visual Sentiment Analysis has applications in the entertainment and security industries, but at the same time they face various problems. Lie detectors loaded with Visual Sentiment Analysis, for example, are useless against well-trained actors and emotionless villains. These expensive machines cannot distinguish between the meanings of various facial expressions that lack true emotion [10]. While there has been some recent successful research and development aimed at recognizing real and fake facial expressions, these are still relatively small-scale and very segmented, such as involving only smiles.

The challenge of speech sentiment analysis: not considering various accents, regional language patterns, personal pronunciation habits, etc.

Lots of non-native speakers keep an accent when speaking a second language. In addition, accents can manifest as transitional tonal, rate, and pause changes that are not specific to the first language. We need a clear explanation for it, otherwise these shifts can lead to misunderstandings of emotions and intentions.

These are issues that need to be focused on overcoming our path to developing better chatbots, intelligent assistants, robotic guides in home and business environments, and ultimately AI that achieves self-awareness, empathy, and a true understanding of human expression.

5. Conclusion

An improved method to represent words is put forward in this thesis and weighted word vectors are generated by integrating sentiment information’s contribution into the traditional TF-IDF algorithm. The experiments of different methods to represent words demonstrate the effectiveness of the method to represent word put forward in this thesis.

The authors put forward a bidirectional CNN-RNN deep model based on attention. ABCDM extracts past and future context by taking the information flow in both directions into account with the utilizing of two independent bidirectional LSTM as well as GRU layers. In this paper, we propose the Deep Contextual Relationship Awareness Network, which allows interaction between subtasks with deep contextual information based on two modules. In particular, we design novel self-supervised strategies for ABSA that have advantages in handling multiple aspects. In the paper, in order to identify opinion words’ sentiment polarity expressed in specific aspect of sentences, a method on the basis of neural networks is proposed. They propose a model which exploits a BI-LSTM to learn the way to represent the characteristics of a sentence, and using a graph convolutional network which operates on the dependency tree of the sentence in a direct manner to further enhance the embeddings. Last but not least, the above is an introduction to several main applications of sentiment analysis and a review of some methods. Sentiment analysis has very broad prospects in security, education, entertainment, finance and other fields, which is inseparable from the upgrading of sentiment analysis technology.

References

[1] Yadav, A., & Vishwakarma, D. K. (2019, December 2). Sentiment analysis using Deep Learning Architectures: A Review - Artificial Intelligence Review. SpringerLink. Retrieved June 5, 2022, from https://link.springer.com/article/10.1007/s10462-019-09794-5

[2] Sentiment analysis of comment texts based on bilstm. IEEE Xplore. (n.d.). Retrieved June 5, 2022, from https://ieeexplore.ieee.org/abstract/document/8684825

[3] Balaganur, S. (2021, November 17). 10 popular datasets for sentiment analysis. Analytics India Magazine. Retrieved June 5, 2022, from https://analyticsindiamag.com/10-popular-datasets-for-sentiment-analysis/

[4] Gómez-Ríos, A., Tabik, S., Luengo, J., Shihavuddin, A. S. M., Krawczyk, B., & Herrera, F. (2018, October 6). Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation. Expert Systems with Applications. Retrieved June 5, 2022, from https://www.sciencedirect.com/science/article/pii/S0957417418306523

[5] Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2018, October 6). Deep learning for aspect-based sentiment analysis: A comparative review. Expert Systems with Applications. Retrieved June 5, 2022, from https://www.sciencedirect.com/science/article/pii/S0957417418306456

[6] Itani, S., Lecron, F., & Fortemps, P. (2018, October 2). Specifics of medical data mining for diagnosis aid: A survey. Expert Systems with Applications. Retrieved June 5, 2022, from https://www.sciencedirect.com/science/article/pii/S0957417418306389

[7] Xiao, Y., Li, C., Thürer, M., Liu, Y., & Qu, T. (2022, May 6). User preference mining based on fine-grained sentiment analysis. Journal of Retailing and Consumer Services. Retrieved June 5, 2022, from https://www.sciencedirect.com/science/article/pii/S0969698922001060

[8] Zhang, Lei, Shuai Wang, and Bing Liu. "Deep learning for sentiment analysis: A survey." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8.4 (2018): e1253.

[9] Hussein, Doaa Mohey El-Din Mohamed. "A survey on sentiment analysis challenges." Journal of King Saud University-Engineering Sciences 30.4 (2018): 330-338.

[10] Liu, Bing. "Sentiment analysis and subjectivity." Handbook of natural language processing 2.2010 (2010): 627-666.