Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/35/20230350

# Review on natural language processing models

**LI YIKUN**

Mathematics Department, University College London, London, UK

zcahibd@ucl.ac.uk

**Abstract.** Accessing information has grown simpler as a result of the internet's expanding use and the arrival of the big data era. Compared to traditional approaches, employing NLP for information condensation and amalgamation proves to be a highly effective method. This article focuses primarily on the sentiment analysis aspect of NLP, offering a comprehensive exploration of two deep learning models: BERT and CNN. It delves into the intricacies of their principles, analyzes their respective strengths and weaknesses, and proposes potential avenues for enhancement. By delving into these models, Researchers and practitioners can obtain a better understanding of sentiment analysis and its applications in diverse fields.

**Keywords:** natural language processing, sentiment analysis, deep learning.

## 1. Introduction

Because of technological advancements, information spreads quickly across the Internet, making network information, news, and public opinion analysis more useful. Unlike the traditional method of reading news manually, using the NLP (Natural Language Processing) model is a more efficient choice, which help to allocate the information in a faster and more accurate way.

NLP, or Natural Language Processing, is a branch of artificial intelligence that seeks to equip machines with the capability to understand, analyze, and generate human language.One of the useful techniques of NLP is the Sentiment Analysis.

Sentiment Analysis, is aimed to analyse sentiment and opinions expressed in economic news articles, social media data, or on-line forums. It will automatically classify the sentiment as as positive, negative, or neutral. The sentiment has been demonstrated by the existing study that the economic news has a relation to the stock price. It can greatly enhance stock price prediction models and make them more accurate.

Sentiment analysis, a field of research within natural language processing, originated from the concept of "affective computing" proposed by Rosalind et al. [1] in 1997. This concept serves as the foundation for sentiment analysis, which involves the analysis and understanding of emotions and sentiments expressed in text data.With the advent of machine learning, there has been more systematic research on text sentiment analysis. In 2002, Pang et al. In the realm of sentiment analysis for movie reviews, [2] presented the introduction of supervised machine learning techniques. They utilized models like Naive Bayes and Support Vector Machines to accomplish this task. In that year, Turney [3] suggested the use of unsupervised sentiment lexicons for analyzing sentiment in online text. In the 21st century, machine learning-based text sentiment classification has gradually become mainstream. Many scholars both internationally and domestically have conducted research on sentiment analysis using

DOI: 10.54254/2755-2721/35/20230350

these methods. Machine learning techniques can be primarily classified into supervised learning and unsupervised learning. Supervised learning is commonly used for sentiment classification tasks. In 2006, Bengio et al. leveraged a three-layer neural network to train language models, successfully obtaining distributed representations of words within a vector space. This groundbreaking approach significantly reduced the vector dimensionality compared to conventional representations. In 2008, Collobert et al. introduced the utilization of neural networks as a proposed approach to automatically learn vector representations of vocabulary. In 2012, Huang et al. [4] believed that localized information of the context cannot fully capture the semantic information of the middle word, so they used global information from the entire text to assist the local information and used multiple word vectors to represent polysemous words. In 2013, Socher et al. used a deep learning method for sentence-level and phrase-level sentiment classification, known as Recursive Neural Tensor Networks. In 2014, Kim revolutionized natural language processing tasks by introducing Convolutional Neural Networks (CNNs) specifically to classify text. Through the incorporation of pre-trained word embeddings, this groundbreaking approach yielded outstanding results and significantly improved the accuracy of text classification tasks. In 2015, Tang et al. [5] utilized CNNs or LSTMs to capture sentence representations and employed GRUs to know document representations. In the year 2018, a team of researchers including Wang Hongsheng and Jin Xiangyu performed sentiment analysis on Chinese e-commerce comments. Their study involved the application of deep learning techniques and advancements in sentiment word embeddings [6]. Li Rong, Liu Shaohui, and others applied a method that combines SVM and RNN for Chinese ambiguous word research, providing important references for the development of Chinese sentiment analysis [7]. In their study, Ramanathan et al. introduced a robust sentiment analysis technique that relied on a combination of convolutional and recurrent neural networks. This approach significantly enhanced the accuracy of sentiment analysis, particularly in the context of longer texts [8]. BERT, proposed by Google in 2018, has delivered outstanding performance across diverse natural language processing tasks, showcasing remarkable results and has achieved the best results on the GLUE dataset.

Deep learning, in its current research state, has demonstrated that constructing models with multiple hidden layers can lead to automatic learning of more valuable features from extensive unsupervised training data. This process enhances the models' ability to generalize, allowing them to better understand complex patterns and improve performance on various tasks. Based on the continuous optimization and iteration of deep neural network models, the accuracy of model results continues to improve, leading to better applications in practical scenarios.

## 2. Dataset

The recommanded dataset is the dataset used in the “Negative Financial Information and Subject Determination” competition held by the China Computer Federation (CCF) in 2019. The dataset consists of 10,000 training samples and 10,000 testing samples. It primarily consists of financial web text (titles and content) and lists of entities in the text, as shown in Table 1.

**Table 1.** Dataset .

| **Field info** | Type | Description |
| --- | --- | --- |
| id | String | Data id |
| title | String | Text title |
| text | String | Text content |
| entity | String | Given entities |
| negative | String | Does the text contain negative information |
| Key_entity | String | Key entity |

DOI: 10.54254/2755-2721/35/20230350

## 3. BERT

### *3.1. Related work*

BERT (Bidirectional Encoder Representations from Transformers) was developed by research scientists Jacob Devlin and his team at Google in 2018.[10] They introduced a novel architecture for pre-training language models that could capture rich sentence representations.

The input representations for BERT are the Token Embeddings, Segment Embeddings and the Position Embeddings.

Token Embeddings: Each input token in BERT is initially represented as a word piece or subword. These subwords are typically generated using a technique called WordPiece tokenization. Each subword token is then mapped to a fixed-size vector called a token embedding. These embeddings capture the semantic meaning of the subwords and provide the initial representation of the tokens in BERT.

Segment Embeddings: BERT can accept input text with multiple segments or sentences. To differentiate between these segments, BERT uses segment embeddings. Each input token is associated with a segment embedding vector, indicating the segment or sentence it belongs to. Segment embeddings help the model understand the relationships between tokens from different segments. For example, in the question-answering task, BERT can distinguish between the question and the answer segments. BERT learns sentence representations through two stages of training.

Position Embeddings: Position embeddings are used to capture the sequential order of the tokens in the input sequence.

BERT learns position embeddings to understand the position-dependent information of the tokens.

By adding position embeddings to the token and segment embeddings, BERT is able to encode both the meaning of individual words and the relative positions of words in the input sequence.

The Pre-trained process for BERT are the two steps:

Masked Language Model (MLM): During this step, BERT randomly masks a certain percentage (typically 15%) of the input tokens in each input sequence. The masked tokens are then predicted by the model based on the surrounding context. By predicting the masked tokens, BERT learns the relationships between words and captures contextual information.

Next Sentence Prediction (NSP): In this step, BERT takes pairs of sentences as input. BERT predicts if the second sentence follows the first sentence in the original text or not, evaluating their logical coherence. This objective enables BERT to grasp the semantic relationships between sentences and capture the overall coherence of text.

These two steps together provide comprehensive language modeling capabilities to BERT. The MLM objective helps BERT learn fine-grained representations of words, while the NSP objective helps BERT capture the relationships between sentences.

After pre-training on a large corpus, Through fine-tuning, BERT can be tailored to specific tasks by leveraging labeled data. During this process, BERT is trained on task-specific data while preserving the majority of the pre-trained weights. This allows BERT to leverage the learned knowledge from pre-training while adapting to the specific task requirements.

### *3.2. Application*

Mr. Zhu and his team developed a Chinese financial text sentiment analysis model based on BERT[9]. In Chinese NLP, the original BERT method splits words into individual characters, causing issues in learning semantic information. To address this, word-based or sub-word tokenization techniques can be used to handle Chinese text. Additionally, special attention should be given to the masking process during training to ensure effective learning of semantic information.

To overcome this limitation, they adopted a financial domain-specific BERT pre-training model with full-word coverage. This means that when some characters within a financial-related word are masked during training, the other characters belonging to the same word will also be masked accordingly. This way, the pre-trained model becomes more suitable and adaptable for tasks in the financial area.

DOI: 10.54254/2755-2721/35/20230350

They constructed a financial domain-specific lexicon on top of the 'GouXibo cell library' and web crawlers to collect financial-related news, comments and other textual information as the pre-training corpus. The original BERT model was then subjected to the financial-related-specific full-word coverage pre-training process.

Through examples, the pre-training method enables the extraction of more domain-specific financial information and effectively tackles challenges related to semantic ambiguity and sparse key features in financial domain comments. For downstream sentiment analysis tasks, a Softmax layer is included in the output layer for the prediction. The overall structure of the model is illustrated in Figure 1.

When comments and financial entities are input into the BERT pre-training model based on full-word coverage in the financial domain, the model's computation formula is as follows:

$$HModel = BERT(T, E) \tag{1}$$

$$Tmodel = MLP(HModel) \tag{2}$$

$$Label = Softmax(Tmodel) \tag{3}$$

In order to improve sentiment extraction and analysis of financial user comments, and overcome the difficulties in feature extraction and the relatively ambiguous boundaries in financial texts, the author presents an enhanced feature representation model.

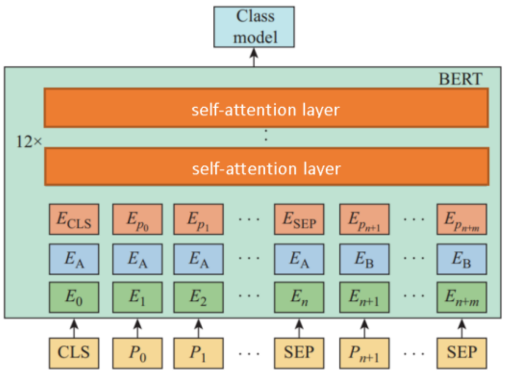


**Figure 1.** Whole word masking based on the BERT model in the financial field.

During the training process of BERT pre-training models, only the relationships between words are considered, however, this model ignores the part-of-speech of the words. The author used the tools of 'jieba fenci' to classify words into noun, verb , adjective and etc.

In additon, since the model cannot learn the continuous features effectively, bidirectional LSTM units are incorporated into the model to address the issue of weak sequential feature learning within sentences.

Table 3 presents the comparison results of different models. It can be observed that the proposed final model in this work achieves significantly improved performance on evaluation metrics Fs1, F1e, and F1, with scores of 86.38, 87.28, and 86.92, respectively. Compared to the original BERT model, it shows substantial improvement, and it also outperforms the other two models (4.6 points higher in F1 compared to XLNet model and 8.8 points higher compared to ERNIE model).

The results show that the proposed sentiment analysis model, leveraging full word coverage and enhanced features in the financial domain, performs precise analysis on financial comments. It demonstrates advantages in sentiment extraction and achieves superior results in financial sentiment analysis tasks.

DOI: 10.54254/2755-2721/35/20230350

**Table 3.** Comparison results between different models (%).

| Model Name | F1S | F1E | F1 |
|---|---|---|---|
| BERT | 78.32 | 82.42 | 80.78 |
| XLNet | 81.7 | 82.68 | 82.29 |
| ERNIE | 76.38 | 79.48 | 78.10 |
| Final Model | 86.38 | 87.28 | 86.92 |

## 4. CNN

### *4.1. Related work*

Convolutional Neural Networks (CNNs) are a unique variant of artificial neural networks that is inspired by the findings of visual neuroscience. In the field of vision neuroscience, Scientists have observed that numerous neurons in the visual cortex possess small local receptive fields, indicating their responsiveness to stimuli confined within a specific region. Some neurons respond only to horizontal lines, some respond to lines in other directions, and some neurons have larger receptive fields. Higher-level neurons derive their stimuli from the reactions of adjacent lower-level neurons. Different receptive fields can overlap and together cover the entire visual field. Building upon this understanding, CNNs have been developed through continuous efforts.

The utilization of convolutional operations is the primary distinguishing feature of CNNs. Convolutional operations involve the extraction of local features from images using convolution kernels, which generate individual neurons. These neurons are then connected in deeper layers, forming the structure of a convolutional neural network. In the following sections, we will delve into the principles, processes, and implementation methods of convolutional operations, starting with the fundamental concepts, and explore them further with practical examples.

The mathematical principle of convolution involves transforming two functions, f(x) and g(x), into another function, h(x), through a specific operation. In CNNs, convolution is primarily used for feature extraction in image processing. In convolution operations, f(x) typically represents the input image, g(x) represents the convolution kernel (also known as a filter), and h(x) represents the convolution result.

The mathematical expression is as follows:

$$h(x) = (f * g)(x) = \int f(x')g(x - x')dx' \tag{4}$$

In the above expression, * denotes the convolution operation, x’ and x represent variables, and dx’ and dx represent infinitesimal increments. From the equation, we can understand that convolution operations involve calculating the integral between f(x) and g(x), and transforming it to obtain the convolution result, h(x).

Original Convolution Construction Process

Input image: First, the input image needs to be represented as a matrix, where each element represents a pixel in the image.

Convolution kernel: Next, a convolution kernel, represented by a small matrix, needs to be defined. The kernel slides across the input image and performs convolution operations at each position.

Convolution operation: The convolution kernel is then used to perform convolution operations on the input image. Specifically, the kernel starts from the top-left corner of the input image and slides across, performing convolution operations at each position to generate a new matrix.

Feature map: Finally, the result of the convolution operation is represented as a feature map. A feature map is a matrix where each element represents a specific feature in the input image.

DOI: 10.54254/2755-2721/35/20230350

### *4.2. Pre-Trained*

The author [12] first use the Word2Vec model to generate word vectors from Google news dataset as the input to the CNN.

Then, in the given context, a window of words is considered and the embedding input vectors within that window are concatenated. The resulting concatenated vector, represented as xi ∈ R(kd), is then subjected to a dot product operation with the weight vector U. Subsequently, a non-linear activation function g is applied to the resulting scalar value, ri, for each window. To enhance the model's capabilities, multiple filters (u1, u2, …, ul) can be employed, and they can be conveniently expressed as a matrix U for matrix-vector multiplication. Additionally, a bias term b is introduced to the equation, allowing further flexibility in the calculations. Thus, ri = g(xi • U + b), with ri belonging to RI, xi belonging to R(k.d*l)" After the convolutional filters are applied to the word vectors, the text processing involves utilizing multiple channels. Each channel corresponds to a different linguistic feature. For example, one channel represents the words themselves, another channel represents the corresponding POS tags, and the third channel represents the shape of the words.

By performing convolutional operations on each channel individually, the result will be m vectors for each channel. These vectors capture unique learned representations and attributes associated with each linguistic feature within the given context.

$$Pi = words(1{:}m) + POS(1{:}m) + Shape(1{:}m) \tag{5}$$

or by concatenation

$$Pi = [words(1{:}m){:}POS(1{:}m){:}Shape(1{:}m)] \tag{6}$$

### *4.3. Evaluation*

Deep learning methods have proven to be highly effective in text classification tasks. In this case, the proposed model achieves an impressive accuracy of 99.07% when trained on a set of samples. When compared to other classification methods such as LSTM, NB, and ELM, it is evident that the combination of deep learning models, specifically CNN and LSTM, yields superior and more accurate results, particularly when dealing with short sentences.

### *4.4. Conclusion*

The paper discusses two widely used models, BERT and CNN, for sentiment analysis. Previous studies have compared these models and found that CNN is particularly adept at capturing local features within phrases or individual words, making it well-suited for shorter sentences. Conversely, BERT excels in handling longer sentences by leveraging its contextual understanding and ability to capture word dependencies. Thus, the choice between BERT and CNN for sentiment analysis depends on the length and complexity of the input sentences, as each model offers specific strengths in different scenarios.

## 5. Conclusion

NLP has made great progress in recent years largely driven by the widespread adoption of deep learning techniques and the availability of vast amounts of data. These developments have enabled notable progress in various NLP tasks such as machine translation, sentiment analysis, and question answering. However, challenges persist in improving semantic understanding and ensuring the quality of text generation. To overcome these challenges, future research in NLP will undoubtedly emphasize enhancing models' understanding of semantics. This entails developing techniques that enable models to comprehend and capture the true meaning and intentions behind human language, allowing for more precise and realistic language processing. In addition to semantic understanding, ensuring the quality of text generation is also a vital area of focus. Text generation systems need to generate coherent, fluent, and contextually relevant outputs that meet users' expectations. By tackling these challenges and advancing the understanding of semantics, the field of NLP will continue to evolve, leading to more accurate, robust, and reliable natural language processing systems.

## References

[1] R.W. Picard's book "Affective Computing" was published by MIT Press in 1997 and covers topics from pages 6 to 9.

[2] B. Pang, L. Lee, and V. Shivakumar presented a research paper titled "Thumbs up? Sentiment Classification Using Machine Learning Techniques" at the Conference on Empirical Methods in Natural Language Processing held in Philadelphia, USA. The paper was published by ACM Press in 2002, and it can be found on pages 79 to 86.

[3] Turney authored a paper titled "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," which was published by ACM Press in 2002. The paper was presented at the 40th Annual Association for Computational Linguistics conference in Philadelphia, USA, and can be found on pages 417 to 424.Hu Bingling. A Review of the Application of Sentiment Analysis in the Stock Market [J]. Journal of Statistics and Applications, 2021, 10(6): 957-962. https://doi.org/10.12677/SA.2021.106100

[4] Huang EH, Socher R, Manning CD, et al. Improving word representations via global context and multiple word prototypes[C]. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 2012(①): 873-882.

[5] Tang D, Qin B and Liu T. Document Modeling with Gated Recurrent Neural Network sentiment Classification[C]. Conference on Empirical Methods in Natural Language Processing, 2015: 1422-1432.

[6] Wang Hongsheng, Jin Xiangyu. Chinese e-commerce comment sentiment analysis based on deep learning[J]. Information Communication, 2018(3): 51-53.

[7] Li Rong, Liu Shaohui, Ye Shiwei, et al. Chinese intersectional ambiguity segmentation method based on SVM and KNN combined[J]. Journal of Chinese Information Processing, 2001, 15(6): 13-18.

[8] Vallikannu Ramanathan, Meyyappan T. Prediction of Individuals Character in Social Media Using Contextual Semantic Sentiment Analysis[J]. Mobile Networks and Applications: The Journal of SPECIAL ISSUES on Mobility of Systems, Users, Data and Computing, 2019, 24(4).

[9] ZHU He, LU Xiaofeng, XUE Lei. Emotional analysis model of financial text based on the BERT[J]. Journal of Shanghai University(Natural Science Edition), 2023, 29(1): 118-128.

[10] DEVLIN J, CHANG M-W, LEE K, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J/OL]. CoRR, 2018, abs/1810.04805.

[11] Kim Y . Convolutional Neural Networks for Sentence Classification[J]. Eprint Arxiv, 2014.

[12] Amit Kumar Sharma, Sandeep Chaurasia, Devesh Kumar Srivastava, Sentimental Short Sentences Classification by Using CNN Deep Learning Model with Fine Tuned Word2Vec, Procedia Computer Science,Volume 167,2020,Pages 1139-1147,SSN 1877-0509, https://doi.org/10.1016/j.procs.2020.03.416. (https://www.sciencedirect.com/science/article/pii/S1877050920308826)