Proceedings of CONF-MLA 2025 Symposium: Applied Artificial Intelligence Research
DOI: 10.54254/2755-2721/2025.BJ27198

The Application of Deep Learning in Text Content Mining and Sentiment Analysis

Siyuan Lu

School of Computer Science, Sun Yat-sen University, Guangzhou, China
lusy53@mail12.sysu.edu.cn

Abstract. With the exponential growth of global data volume in the digital era and the increasing proportion of unstructured data—such as user reviews, industry reports, and social media posts—enterprises and institutions are in urgent need of filtering reliable decision-making bases from massive text content to drive strategic planning and operational optimization. Traditional manual processing methods, however, are no longer viable due to their low efficiency, high labor costs, and susceptibility to subjective biases, especially when facing data scales of millions or even billions. Breakthroughs in natural language processing (NLP) technology have significantly enhanced machines' ability to understand text semantics, moving beyond simple keyword recognition to grasp contextual logic and implicit meaning. Meanwhile, the integration of multi-source data (e.g., text, user behavior, and temporal-spatial data) with text mining has broken information silos, enabling more comprehensive and contextualized content analysis. This thesis mainly studies the role of deep learning in text content mining and sentiment analysis. It finds that innovative designs —including deep learning models like BERT and graph neural networks, as well as hierarchical attention mechanisms—have notably strengthened semantic understanding and emotion capture capabilities. Additionally, the fine-tuning of pre-trained models has opened up new avenues for tailored applications in specific fields such as healthcare and finance.

Keywords: BERT, Attention mechanism, Neural network, Deep learning

1. Introduction

In the current information explosion era, the Internet and social media have produced massive text data. As key technologies for extracting valuable insights from these data, text content mining and sentiment analysis have gained wide attention from academia and industry. Deep learning, especially pre-trained language models like BERT and GPT, has greatly advanced this field with its excellent semantic understanding and contextual representation capabilities. Nevertheless, despite remarkable achievements, existing research still faces challenges: general models are inadequately adaptable to specific domains (e.g., healthcare, law), their ability to identify complex emotional scenarios (such as sarcasm or multimodal expressions) is limited, and models have efficiency and interpretability issues in low-resource environments. This study aims to systematically sort out the cutting-edge applications and innovations of deep learning in addressing these challenges. It will explore three core aspects: analyzing pre-trained model fine-tuning, lightweight design and neural network

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.BJ27198

structure innovations; conducting in-depth research on fine-grained, multi-type and cross-domain sentiment analysis technologies; and covering text mining applications like topic modeling and information extraction, as well as interdisciplinary issues such as few-shot learning, interpretability and ethical bias. Through a comprehensive review of advanced methods, this paper intends to provide researchers and practitioners with a clear technical roadmap and practical guide, promoting the application of intelligent text analysis in scientific and technological communication, social public opinion monitoring, and offering references for its future development.

2. Basic algorithms and models

2.1. BERT model

BERT(Bidirectional Encoder Representations from Transformers) is a pre-trained language model proposed by Google AI Research in 2018, based on the Encoder structure of Transformer, which is of milestone significance in the field of natural language processing (NLP) [1].

The input consists of Word Embedding, Segment Embedding(used to distinguish sentence pairs), and Position Embedding(used to mark the position information of words). The final embedding vector is the sum of the three.

With the multi-layer bidirectional Transformer as the core, the comprehensive awareness of text context is realized through multi-layer encoding and information transmission. Different from traditional static word embedding (such as Word2Vec), BERT can dynamically adjust the word vector according to the context, effectively solve the "polysemy" problem, so as to capture richer semantic features.

The specific pre-training tasks are as follows:

First, Vit is used to encode the image and obtain the feature vector V1: $\mathrm{ViViT}(\mathrm{img}_{\mathrm{path}})$.

$\mathrm{img}_{\mathrm{path}}$ is the input image path. Then the image blocks are divided into tokens, and the features are extracted by Transformer. Then, the text feature vector is obtained by embedding the prompt $\mathrm{V_T} = \mathrm{Embedding}\,(\mathrm{prompt})$. Next, the image features and text features are concatenated to obtain the fusion features $\mathrm{V_{1,T}} = \mathrm{Concat}\,(\mathrm{V_1}, \mathrm{V_T})$. Finally, the image text description is output by Qwen2 large language model $\mathrm{D_{I,T}} = \mathrm{Qwen2}(\mathrm{V_{I,T}})$, In this way, the multimodal task is transformed into a text processing task.

2.2. Innovations in neural network architecture

CGCBVul model has significant innovation in neural network structure [2]. The first is the graph feature extraction module. In terms of multi-channel feature learning, the centrality analysis (betweenness centrality, eigenvector centrality) is introduced to expand the node feature matrix of the code attribute graph into a multi-channel feature matrix, and the Graph Attention Network with Shared Weights (GATv2) is used to learn multi-dimensional graph features. GATv2 solves the problem that the attention coefficient is independent of the query node in traditional GAT through the dynamic attention mechanism, and its feature update formula is

$$\mathrm{h_i} = \sigma(\sum j \in N\,(\mathrm{i})\alpha_{\mathrm{i,j}}, W\mathrm{h_j}) \tag{1}$$

DOI: 10.54254/2755-2721/2025.BJ27198

Wh_j is the new feature of node. $\alpha_{i,j}$ is the attention coefficient between nodes i and j, which is calculated by the following formula:

$$\alpha_{i,j} = \frac{\exp\left(a^T \text{LeakyReLU}(W*[h_i h_j])\right)}{\sum_{k \in N(i)} \exp(a^T \text{LeakyReLU}(W*[h_i h_j])} \tag{2}$$

In this way, the extraction ability of structural semantics is improved.

Then multi-scale 2D_CNN was used to further extract the multi-channel features of GATv2 output

$$H_g = FC\left(\|_{k=1}^{K} MaxPool\left(\sigma\left(\text{conv2D}_k\left(H_g\right)\right)\right)\right) \tag{3}$$

Through convolution kernels of different sizes, the short and long distance dependencies of nodes are captured at the same time, and the Max pooling and fully connected layer are combined to realize the effective fusion of multi-channel features.

For the sequence feature extraction module, the dual-scale 1D-CNN is used to capture multi-scale sequence dependencies, and the short-range and long-range dependencies are fused by an averaging operation. The formula is

$$H_c = AVG(Conv1D1\left(H\right), Conv1D2\left(H\right)) \tag{4}$$

Here, H_c is the sequence feature that is input to the BiLSTM layer. The small convolution kernel captures the short distance, and the large convolution kernel captures the production distance, which enhances the capture of different scale sequence semantics.

In addition, the global sequence features are innovatively extracted by BiLSTM:

$$H_b = BiLSTM(H_c) \tag{5}$$

Where H_b is the enhanced sequence feature.

The attention mechanism is combined to dynamically allocate weights to obtain better sequence features

$$H_s = H_b * softmax(\tanh\left(H_c\right)W) \tag{6}$$

This design makes the model focus on the sequence fragments critical to vulnerability detection, and improves the targeted extraction of sequence semantics.

2.3. Hierarchical attention mechanism

The hierarchical attention mechanism is used to capture the emotional dependence between words and sentences, especially for the emotional association between sentences in long texts, which can

DOI: 10.54254/2755-2721/2025.BJ27198

distinguish the contribution of different sentences to the overall sentiment of the text. The specific design is as follows:

Firstly, the article level architecture is divided into word level and sentence level architecture. For the word-level architecture, the convolutional attention mechanism is used to capture the emotional dependence between words in the sentence and generate the sentence representation.

For each statement s_d in the discourse(From the sequence of words $\{ x_1^d, \ldots, x_T^d \}$), which is first processed using a sentence-level convolutional Attention network (CAN):

The bidirectional LSTM is used to encode the context of words to obtain the bidirectional semantic representation of words. And concatenate the forward and reverse semantics to obtain the comprehensive representation $h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t}$. Then, the convolutional attention mechanism is used to calculate the attention weight α_t of the word, which reflects the importance of the word for the sentiment of the sentence. The formula refers to the smoothing and normalization process of the convolutional attention.

$$c_t^{'} = \frac{1}{l} \sum_{t'=t-l+1}^{t} c_{t',t'+l-1} \tag{7}$$

$$\alpha_t = \sigma\left(c_t^{'}\right) = \frac{\exp\left(c_t^{'}\right)}{\sum_{t'} \exp\left(c_t^{'}\right)} \tag{8}$$

Where $c_{t',t'+l-1}$ is the similarity between the convolution kernel and the continuous word segment, and l is the length of the convolution kernel.

The statement can be obtained by a weighted summation as

$$r_d = \frac{1}{T} \sum_{t=1}^{T} \alpha_t h_t \tag{9}$$

For the sentence-level architecture, the document-level convolutional attention mechanism is used to capture the emotional dependence between sentences in the document and generate the document representation. The specific formula is roughly the same as the word-level, but the scale is larger, which is not described here.

The core role of the hierarchical attention mechanism is to distinguish the importance of the text. For example, in the movie review, the attention mechanism will assign a higher weight to the former according to the fact that the "plot delay" has a greater negative impact than the "compact structure".

3. Segmentation scenarios of emotions

3.1. Fine-grained sentiment analysis

Compared with coarse-grained analysis, fine-grained sentiment analysis can dig deeper into the sentiment details in the text [3]. It can be divided into two categories: One is to group emotions into specific categories (e.g., joy, anger, sadness, etc.). The second is to classify by specific attributes

DOI: 10.54254/2755-2721/2025.BJ27198

(such as price, quality of a product, or availability, timeliness of scientific datasets, etc.) and analyze the polarity of opinions and sentiments on each attribute.

However, fine-grained sentiment analysis faces two major challenges: the lack of scientific data sets leads to poor generalization ability, and if the generative model is used to deal with attribute-level analysis, it may produce multiple reasonable triple results due to the dependence on maximum likelihood estimation, so the model cannot directly determine the optimal solution.

Therefore, Gao Yuwei proposed the fine-grained sentiment analysis method EOG-ALSA [4]. The main steps are as follows:

Firstly, the attribute, opinion, and sentiment polarity are respectively expressed as [A][O][S], and the suggestion template is constructed through the arrangement and combination of the three to enhance the model's ability to adapt to different element orders, and the suggestion sentence is added before and after the original comment to clearly guide the triplet extraction task of the generation model.

Based on the pre-trained generative model T5, the generation is constantly fine-tuned, and its generation ability is as follows:

$$P(\mathrm{yx}) = \prod_{t=1}^{T} \mathrm{P}(\mathrm{y_t}|\mathrm{y_1}, \ldots, \mathrm{y_{t-1}}, \mathrm{x}) \tag{10}$$

Where x is the input sequence (comments with A hint template), y is the sequence of generated triples (e.g. "[A] timeliness [O] is good [S] positive"), and T is the length of the generated sequence.

For the problem that the optimal solution cannot be determined, A voting mechanism is designed to extract [A][O][S] from the designed triple and combine them into a new sentence. Sentent-BERT is used to encode the original comment and the new sentence into a semantic vector, and the semantic similarity is calculated by the following formula [5].

$$Sim_{\mathrm{score}} = Cosine_{\mathrm{sim(D,T)}} \tag{11}$$

Where D is the semantic vector of the original review and T is the semantic vector of the new sentence.

Finally, the triple with the highest similarity is selected as the final result output.

3.2. Analysis of specific emotion types

In social media, people often express emotions through sarcasm, humor, metaphor and other rhetoric, and sarcasm detection is an important part of it.

In the existing research, the UCCA-GAT model is used by transforming the semantic structure into a computable matrix representation [6]:

The adjacency matrix $\mathrm{A} \in \mathbb{R}^{n*n}$ is used to describe the node connection relationship, n is the number of nodes, the feature matrix $\mathrm{X} \in \mathbb{R}^{n*k}$ is used to generate the node embedding through the pre-trained language model BRET, and k is the embedding dimension.

The model learns node weights through a graph attention mechanism, and the hidden layer is calculated as follows:

DOI: 10.54254/2755-2721/2025.BJ27198

$$H^{i+1} = \sigma\left(A \cdot H^{i} \cdot W^{i}\right) \tag{12}$$

Where H^{i} is the ith layer feature matrix, W^{i} is the weight matrix, σ is the $ReLU$ activation function, and the final classification result $Z = \mathrm{softmax}(H^{l})$ (is output by $softmax$, H^{l} is the final feature matrix, Z is the probability distribution of sarcasm/non-sarcasm. In this way, NIP can be used to break through the literal meaning and combine context and commonsense reasoning to obtain the correct semantics.

3.3. Domain-specific sentiment analysis

In the aspect of news public opinion processing, text analysis is more important. Firstly, it preprocessed the public opinion data, used $jieba$ library to convert the text into analyzable components, and applied $SnowNIP$ algorithm to evaluate the sentiment. According to the threshold of 0.4 and 0.6, the emotion is divided into negative, neutral and positive emotions, and then the heat and emotion are calculated according to the following formula [7]:

$$Heat = \sum_{i \in D(t)} \left(x_i^{pos} + x_i^{neu} + x_i^{neg}\right) \tag{13}$$

Where $x_i^{pos}, x_i^{neu}, x_i^{neg}$ are all 0-1 variables, indicating the emotion type of the post.

$$Mood_t = \frac{M_t^{pos} - M_t^{neg}}{M_t^{neg} + M_t^{pos}} \tag{14}$$

Where M_t^{pos}, M_t^{neg} are the total number of public opinion posts of positive and negative sentiment types in a day, respectively, which are used to measure the relative bullish sentiment of the public on the news.

When the Heat exceeds a certain threshold, the monitoring upgrade is triggered. Combined with the "mutation point" of the emotion, the timely intervention and guidance can be obtained, and the analysis results with more accurate semantic understanding and stronger generalization ability can be obtained.

4. The technical application of text mining content

4.1. Semantic annotation technology

Semantic Role labeling (SRL) is the core task of semantic annotation technology [8]. The goal is to identify the predicate in the sentence (the core word of the action or relation) and determine the argument associated with the predicate (the entity or component involved in the action). In this paper, the relationship between predicate and argument is transformed into graph structure (boundary connection or full connection) by designing four graph annotation modes: BES, BE, BIES and BIE [9]. For example, all predicates are connected by "Root" pseudo-nodes, and argument boundaries and roles are marked with edge labels such as B-r and E-r. At the same time, a connective-relation mapping table is designed to filter unambiguous connectives, and the ambiguous

DOI: 10.54254/2755-2721/2025.BJ27198

connectives in the training set are processed by a coarse/fine-grained replacement strategy. Then, the flattened linearization (taking the role sequence of the first predicate) or nested linearization is transformed into sequence features to enhance the input semantic representation [10].

The use of this technology can not only accurately identify the semantic role relationship in the text, clearly present the shallow semantic structure of the sentence, but also improve the performance of natural language processing downstream tasks, providing semantic support for machine translation, information extraction, question answering systems, discourse relation recognition and other tasks.

4.2. Text generation and summarization

Automatic text generation summarization aims to preserve the core semantics by compressing the text, which is divided into extractive summarization and generative summarization [11]. It can model text semantics through a neural network based on deep learning method, and significantly improve the quality of summary. This paper mainly introduces extractive summarization, whose core logic is to directly select key sentences or phrases from the original text and concatenate them to form a summary without generating new text content.

First, the SummaRuNNer model is used to generate the average pooling of sentences via word and sentence RNN modeling [12]:

$$\mathrm{s_i} = avg(\mathrm{w_{i1}}, \mathrm{w_{i2}}, \ldots, \mathrm{w_{ik}}) \tag{15}$$

Here, $\mathrm{w_{ij}}$ is the $\mathrm{j^{th}}$ word embedding in sentence i and $\mathrm{s_i}$ is the sentence vector.

Next, the sentence-word bipartite graph is constructed, and the edge weights are defined by TF-IDF weights:

$$w\left(\mathrm{e_{w,s}}\right) = TF - IDF(w, s) \tag{16}$$

Where $\mathrm{e_{w,s}}$ is the edge between word w and sentence s. The node importance is updated through the graph attention network. Finally, the sentence embedding is generated by BERT. Use $\mathrm{score}\left(\mathrm{s_i}\right) = \mathrm{MLP}\left(\mathrm{BERT}\left(\mathrm{s_i}\right)\right)$ to calculate the sentence score, and the high-score sentences are selected to form the summary.

5. Conclusion

This paper focuses on the application of deep learning in text content mining and sentiment analysis, focusing on its practical application in the basic model, sentiment analysis and text mining. Research shows that deep learning models such as BERT and graph neural network improve the ability of semantic understanding and emotion acquisition. Innovative designs such as hierarchical attention mechanisms and neural networks can more accurately capture text context dependencies and details. Fine-tuning of pre-trained models opens up new research avenues for special fields such as news public opinion analysis. However, this research also has some limitations. It has not explored low-resource language scenarios and multi-modal text mining, such as image fusion, audio fusion, etc., and there is insufficient research on the interpretability of complex models when dealing

DOI: 10.54254/2755-2721/2025.BJ27198

with expressions such as sarcasm and metaphor. In the future, the ethical bias problem in sentiment analysis can be solved by constructing a multi-modal fusion framework and optimizing the model's lightweight, so as to enhance the practical application value of deep learning technology in text content mining.

References

[1] Yang Yufei, Qian Yurong, Gong Weijun, et al. Multimodal Sentiment Analysis based on BERT with Large Model Optimization [J/OL]. Applications in Computer Systems
[2] Chen Kejia, Wang Zhenqing, Zhou Xiukao. A Text Classification Model Based on BERT-sPTT [J/OL]. Computer Engineering and Applications
[3] Li Peng-wei, ZHENG Hong, SHAN Rong-sheng. Vulnerability Detection Method Based on Multi-channel Graph Neural Network and CNN-BiLSTM [J/OL]. Journal of East China University of Science and Technology (Natural Science Edition)
[4] Gao Yuwei, Fang Jin, Hu Lianglin, et al. Fine-grained sentiment analysis method for social comments of scientific data sets [J/OL]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)
[5] Wang, Zhen and Li, Honglei. Research on the Construction of Topic Model in the Field of Digital Humanities Based on Sentence-BERT and LDA. Association for Computing Machinery
[6] Necva , Can and Burcu. Semantically Informed Graph Neural Networks for Irony Detection in Turkish. Association for computer Science
[7] Chen Duan, Wang Yining, and Zhang Junyi. Research on public opinion Transmission and capital market Reaction mechanism of Artificial Intelligence Technology Innovation Events: Multi-level network analysis research. Association for computer Science
[8] Yang Haoping, Xia Qingrong, Li Zhenghua, et al. GCN based Heterogeneous syntactic driven Chinese semantic role labeling [J]. Computer Applications and Software, 2007.
[9] Ban Ma Bao, Luo Peng, First Denier, etc. An End-To-End Semantic Role Labeling Method for Tibetan Fragments Based on Graph Parsing [J]. Journal of Beijing University (Natural Science Edition)
[10] Wu Yiheng, Li Junhui, Zhu Muhua. Semantic Role labeling enhanced Implicit Discourse Relation recognition [J]. Journal of Xiamen University (Natural Science Edition)
[11] Qi Qi Li Ge, Schentu, Wang Si ri Gulen. Automatic text summarization based on deep learning: A Survey [J/OL]. Computer Engineering and Applications.
[12] Anh, Bui Thi Mai and My, Nguyen Tra and Trang, Nguyen Thi Thu. Enhanced genetic algorithms for Single document Extraction summarization. Association for computer Science