Available Online: 28 May 2025 DOI: 10.54254/2977-3903/2025.23585

Fine-grained sentiment analysis for social media: from multi-model collaboration to cross-language multimodal analysis

Yuanmiao Dong

Department of Information Technology, College of International Business and Economics, Wuhan, China

3067754214@qq.com

Abstract. With the rapid development and widespread popularity of the Internet, the amount of data in social media and networks is growing exponentially, and sentiment analysis for this huge amount of data is very complex but significant. Fine-grained sentiment analysis has become the choice of researchers when dealing with various sentiment analysis tasks. Different from coarse-grained sentiment analysis, which only focuses on emotional polarity, fine-grained sentiment analysis involves emotional polarity and emotional intensity and recipients, providing more specific information about emotions. This paper aims to provide relevant research methods on fine-grained sentiment analysis and apply them to social network texts to analyze the challenges and solutions. This paper will classify fine-grained sentiment analysis from three methods: rule-based, machine learning and deep learning. This research finds that fine-grained sentiment analysis can not only accurately capture the emotions in the text, but also judge the direction and intensity of emotions, and understand different types of emotions in the text more specifically. This is of great help in dealing with more complex texts, such as social network texts. Combining fine-grained sentiment analysis with various large models can solve many challenges and problems when dealing with social network texts.

Keywords: fine-grained sentiment analysis, deep learning, machine learning

1. Introduction

Fine-grained sentiment analysis refers to not only judging the positive or negative emotional tendency of the text as a whole, but also further deepening into all aspects and elements of the text, accurately identifying and analyzing the specific emotional categories, emotional intensity, emotional objects and other more specific and detailed emotional information contained in the text analysis process. Fine-grained sentiment analysis can be widely used in product and service evaluation analysis, public opinion monitoring and management, user research and market research, as well as personal emotional and mental health. This paper analyzes the core tasks and framework of fine-grained sentiment analysis. There are four main tasks, which are multi-level sentiment classification, sentiment intensity analysis, aspect-level sentiment analysis, and multi-dimensional sentiment analysis. Among them, the framework of aspect-level sentiment analysis is mainly divided into four parts: text representation, feature extraction, sentiment classification, and post-processing. Compared with fine-grained sentiment analysis, coarse-grained sentiment analysis is very limited in extracting emotional information from text. At the same time, fine-grained sentiment analysis is combined with various large models, such as Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), and Self-Attention. Combining the three models with fine-grained sentiment analysis to construct local, temporal, and global features can greatly improve the accuracy of processing complex texts. In addition, the Pre-Trained Language Model (PTLM) is combined with fine-grained sentiment analysis, and is analyzed through three steps: self-supervised learning, transfer learning, and context awareness. The most classic Bidirectional Encoder Representations from Transformers (BERT) model in the PLTM model can dynamically capture text information and improve model processing performance.

2. The core tasks and framework of fine-grained sentiment analysis

Fine-grained sentiment analysis aims to classify the emotions in the text more specifically and meticulously. There are four main tasks: multi-level sentiment classification, sentiment intensity analysis, aspect-level sentiment analysis, and multi-dimensional sentiment analysis. Multi-level sentiment classification is to divide emotions into multiple levels, such as 1 star to 5 stars, which is a more specific quantification of text emotions; emotional intensity analysis is to analyze the strength of emotions, which is

usually divided into very negative, negative, neutral, positive and very positive, mainly to capture the subtle differences of emotions. Aspect-level sentiment analysis is to identify the specific aspects mentioned in the text and analyze the emotional tendencies of all aspects, which can provide more specific emotional feedback expressed by the text; multi-dimensional sentiment analysis is to analyze the emotions of different dimensions in the text, such as happiness, sadness, anger, panic, etc., which will identify more complex emotional expressions in the text.

The framework of fine-grained sentiment analysis is mainly divided into four key parts: text representation, feature extraction, sentiment classification, and post-processing. More easily understood, it is to convert the text into a machine-understandable numerical representation, and then extract the features related to emotion from the text, and then classify the emotion according to the extracted features, and finally optimize and adjust the classification results.

2.1. The difference between coarse-grained and fine-grained sentiment analysis

Coarse-grained sentiment analysis belongs to document-level sentiment analysis, that is, sentiment classification of the entire text or paragraph. It is a relatively broad classification of emotions in the text. It only involves emotional polarity, including positive, negative and neutral, regardless of emotional intensity or specific aspect. It is mainly suitable for tasks that need to quickly judge the overall text emotion. In some cases, coarse-grained sentiment analysis may have only positive and negative categories. Fine-grained sentiment analysis is more detailed and specific for text processing, involving more aspects of the text. It divides emotions into multiple levels, analyzes the strength of emotions, identifies specific aspects of emotional tendencies in the text and so on. For today's various social network texts, such as Weibo comments, buyer evaluation, and public opinion monitoring, fine-grained sentiment analysis is more needed for identification and judgment.

2.2. An important branch of fine-grained sentiment analysis—ABSA

Fine-grained sentiment analysis is a generalized concept that analyzes the emotions covered in the entire text or sentence. Aspect-Based Sentiment Analysis (ABSA) is a specific task of fine-grained sentiment analysis, which focuses more on independent analysis of different aspects of emotions in the text, rather than sentiment classification of the whole text. ABSA is mainly understood from four elements: aspect term refers to the opinion target that clearly appears in a given text, aspect categories defines a unique aspect of an entity, opinion terms is the opinion expressed by the opinion holder on the target, sentiment polarities describe the tendency of emotion in a certain aspect category or aspect. For example, for the sentence “this denim coat is cool”, the aspect term is “denim coat”, the aspect category is “clothes”, the opinion term is “cool”, and the sentiment polarity is positive.

These four elements are mainly applicable to a single Aspect-Based Sentiment Analysis task, that is, to identify each single emotional element in the sentence. Contrary to a single ABSA task involving only a single emotional element, the research on the joint prediction of multiple emotional elements in the text is a composite ABSA task, which may involve a pair, three even four emotional elements. The composite ABSA task is mainly analyzed from three elements: pair extraction, triplet extraction, and quad extraction. When dealing with aspect-level sentiment analysis, one of the operations that usually needs to be performed is to embed pre-trained words, that is, to perform pre-training. At the same time, this also brings a new problem to be solved. A context-free word embedding cannot recognize the complex emotional relationship in the sentence, and the current research shows that the pre-trained language model PLM, such as BERT and RoBERTa can deal with this kind of context-based embedding problem to a certain extent [1].

For example, Professor Li et al. proposed an end-to-end ABSA model based on BERT, which integrates aspect extraction and sentiment classification into a unified framework. BERT is used to encode the input text to generate the context representation of each word, and then BIO (Begin, Inside, Outside) is used to sequence each word to identify aspect words and sentiment polarity, which greatly improves the ability of BERT model to deal with aspect-level sentiment analysis tasks.

3. The fine-grained sentiment analysis method classification

According to technical means, Fine-grained sentiment analysis methods can be divided into three aspects: rules-based, machine learning, and deep learning methods.

3.1. Rule-based approach: sentiment dictionary

Rule-based fine-grained sentiment analysis is a traditional sentiment analysis method. It mainly relies on artificially defined rules, dictionaries and linguistic knowledge to achieve a detailed classification of text sentiment. Its core idea is the sentiment dictionary and rule engine. The sentiment dictionary is a dictionary that contains the emotional polarity and intensity of words. The rule engine handles more complex situations such as negative words, degree adverbs, and context dependence through artificially defined rules. This paper will focus on the analysis of the commonly used dictionary SentiWordNet in sentiment dictionary [2].

In 2006, Professor Esuli and Sebastiani introduced SentiWordNet, a sentiment dictionary based on WordNet, which is mainly used for opinion mining and sentiment analysis. Its main construction method is to use semi-supervised learning method to assign

three scores to each WordNet synonym set: Ogj (s), Pos (s), Neg (s), which represent the objectivity, positive emotion and negative emotion of words respectively, so as to classify them, and then sum up the results of eight ternary classifiers to obtain SentiWordNet scores. The relevant experimental data show that some synonyms containing opinion-related attributes account for 24.63 % of the total WordNet, and the synonym sets of adjectives and adverbs tend to be more subjective.

However, this dictionary lacks the resources to fully manually annotate WordNet, and the accuracy obtained is often only used as a reference, and it is impossible to know exactly the actual accuracy. Therefore, SentiWordNet is very suitable when it is necessary to distinguish between positive, negative and objective emotions. For example, when tracking the emotional tendencies of users on social networking platforms towards a brand, commenting that “the brand's recently released new product makes me feel nothing new”, then “nothing new” is a negative emotion.

3.2. Methods based on traditional machine learning

Unlike rule-based methods, machine learning methods automatically learn emotional features by training models and can handle more complex text sentiment analysis. Feature extraction and classification model is a typical method based on traditional machine learning. Feature extraction is divided into bag-of-words model, TF-IDF and n-gram features. The n-gram feature is to extract continuous n words as features to capture local semantic information. Common classifiers are divided into Naive Bayes, Support Vector Machine, etc.

Professor Pang et al. used three machine learning algorithms: Naive Bayes, Maximum Entropy Classification and Support Vector Machine to test film reviews under the Bag-of-Words model [3]. They constructed data sets with positive, negative and neutral comments, screened out comments whose authors’ scores were represented by stars or values, and studied the influence of different features on the classification effect. The experimental results show that the accuracy of machine learning algorithm in the sentiment classification task is far more than that of the rule-based method. Support vector machine has the highest accuracy among the three classifiers.

Machine learning algorithms can be applied not only to film-related reviews, but also to financial market forecasting. Professors Doriset et al. combine sentiment analysis with financial indicators to evaluate the effectiveness of sentiment analysis in market forecasting [4]. Based on the results, they provide investors with practical and effective trading strategies. Their experiment is to construct data sets from various financial news networks or historical markets. After data preprocessing, a variety of classifiers are used for sentiment analysis, including random forest and LSTM network, and then various index evaluation models are used for performance analysis. The experimental results also show that machine learning algorithms are significantly better than dictionary-based methods in sentiment classification tasks, especially the two machine learning models of random forest and LSTM network.

3.3. Methods based on deep learning

Deep learning is to automatically learn complex patterns and feature representations from a large amount of data by constructing a neural network model with a multi-layer structure. Fine-grained emotional computing based on deep learning is to use deep learning models to deeply analyze the emotional tendencies in the text and mine more accurate emotional information.

3.3.1. Sequence modeling based on a combination of multiple models (CNN+LSTM+Self-Attention)

CNN (Convolutional Neural Network) is a deep learning model that specializes in processing images, audio, and text. It is mainly divided into three parts: the first is the convolutional layer, the pooling layer, and the fully connected layer. The convolution layer is the core of CNN, and the local features between words in the text are extracted through the convolution layer. The pooling layer is mainly used to reduce the feature dimension and calculation amount, and retain important information. Generally, there are two pooling methods: maximum pooling and average pooling. The fully connected layer is to connect the important data in the pooling layer and adjust the classification according to the weight.

In Professor Kim's research, CNN is used to do the task of sentence-level classification [5]. The sentence is represented as the splicing of the word vector. The convolution kernel slides on the sentence, and the word vector is convoluted to generate new features. In the pooling layer, the maximum pooling is taken, and the maximum value in the new feature is selected as the representative to capture the important features in the text. Then the softmax function is calculated in the full connection layer to obtain the probability distribution of different emotions in the text. At the same time, in order to prevent the over-fitting problem, the constraints of relevant model data are made in the penultimate layer. Experiments show that CNN based on pre-trained word vectors can perform well in sentence-level classification tasks, and complex sentiment analysis is not helpless.

CNN is good at capturing local features, LSTM solves the problem of long-distance dependence, and Self-Attention captures the relationship of global context. The next paragraph will specifically discuss the combination of the three models to maximize the advantages of text processing.

At present, the answer given by the latest research is CNN → LSTM → Self-Attention. The CNN layer extracts local features at the word level, LSTM models bidirectional context dependence, and Self-Attention dynamically weights important time steps.

The accuracy of this model in text classification tasks is as high as 93.7 %. First, Professor Kim uses CNN to classify text on MR data sets with an accuracy of 85.0 %, and Professor Yang on Yelp comment data sets. Combining BiLSTM and Attention to achieve 68.2 % F1 value, Professor Wang combined CNN and LSTM models on the AG News dataset to achieve an accuracy of 91.1 % [5-7]. Professor Zhou added Self-Attention mechanism on the basis of CNN-LSTM to reproduce the experiment of AG News and applied it to the Chinese news classification task [8]. The accuracy rate reached 93.2 %. It can be seen that the combination of the three models, through the local-temporal-global three-level feature modeling, greatly improves the performance of complex sequence tasks, and plays an important role in areas requiring high-precision modeling such as medical text analysis and financial time series prediction.

3.3.2. The innovation of pre-trained language model PTLM

The innovation of pre-trained language model has completely changed the field of natural language processing. In simple terms, PLTM is a model that learns the general representation of language by pre-training on large-scale text data. Its core is to learn language rules first and adapt to specific tasks. PTLM has three core points. The first is self-supervised learning, that is, learning directly from the original text without manually labeling data. The second is transfer learning, which transfers the learned general knowledge to downstream tasks, such as text classification. Finally, context awareness, which can dynamically adjust the representation of words according to the context, such as 'Apple' has different meanings in 'Eat Apple' and 'Apple Mobile Phone'.

In the PTLM model, the most dazzling is the BERT model, which learns the deep rules of language through pre-training. At the same time, efficient transfer learning can adapt downstream tasks with a small amount of labeled data, and can be achieved through tool chains such as Hugging Face.

The BERT model is completely based on the bidirectional encoder of Transformer, and obtains the left and right information of each word through Self-Attention at the same time [9]. The Transformer model abandons the loop and convolution structure and implements parallel computing and full-set context modeling based entirely on the self-attention mechanism model [10]. In the pre-training stage, MLM (Maseked Language Model) is used to randomly mask some input words and predict them, and then NSP (Next Sentence Prediction) is used to input sentence pairs to determine whether the sentences constitute a context relationship. In the fine-tuning phase, only the input and output layers of a specific task are added, and [CLS] and [SEP] tags are added. For short text tasks, BERT-base-uncased (English) or BERT-base-Chinese (Chinese) is selected. For long text, Longformer is selected to fine-tune all parameters end-to-end.

The ablation experiments carried out by Devlin et al. The results show that the training BERT model performs better than the one-way model on all tasks [11]. Removing NSP will significantly reduce the performance of the model in dealing with tasks. Bidirectional context modeling solves the problem that traditional models (such as one-way RNN) cannot dynamically capture context, marking that NLP has entered the pre-training era.

4. Challenges and solutions of social network text

Social network text has the characteristics of short text, colloquialism, multilingual mixing, and combination of pictures and texts, including some informal languages, such as network hot words or dialects. There may also be linguistic phenomena such as irony or abbreviation. At this time, it is necessary to understand the implicit emotions in the context, and carry out accurate sentiment analysis through transfer learning and low resource processing.

4.1. Cross-language sentiment analysis: pre-training and low-resource language processing

Generative pre-training has a significant effect on understanding English, but most of the related research is in the single language and mainly focuses on English. In response to this challenge, Professor Lample and Conneau proposed three language modeling goals, namely CLM (Causal Language Modeling), MLM (Masked Language Modeling), and TLM (Translation Language Modeling) [12]. The first two require only monolingual data, and the latter requires parallel data. The Cross-Language Model (XLM) is used for zero-shot cross-language classification, unsupervised and supervised neural machine translation, low-resource language modeling, and unsupervised cross-language word embedding. According to the previous words in the sentence, the probability of the next word is predicted, and some words in the random mask text are predicted to capture two-way context information. Then, by splicing parallel sentences, the words in the source language and the target language are randomly masked, so that the model can represent different languages and improve cross-language learning ability. In the experiment, in the low-resource language modeling task, using data from similar languages (such as Hindi) and distant languages (such as English) can reduce the confusion of the Nepali language model. Therefore, XLM can reduce the English-centered bias and construct a universal cross-language encoder.

4.2. Multimodal applications: BERT and ResNet model combination, OTEModel model innovation

Multimodal sentiment analysis is also the focus of current research, combining the emotional information of text, image, voice and other modal data. In the social network text, there will be a situation where the image and text emotions are consistent or conflicting, the social media image contains irrelevant elements, or the text and image features are inconsistent. To solve this problem, Professor Ren combines BERT and ResNet for image and text sentiment analysis [13]. BERT is used for text processing, ResNet is used for image processing, and five multi-modal models based on BERT and ResNet50 fusion are proposed. The multi-modal public data set MAVA-single training model from Twitter is used for evaluation. Experiments show that. The OTEModel model performs best in various evaluation indicators, and has stronger comprehensive performance in sentiment analysis tasks, and can more accurately identify the emotional tendency of pictures and texts.

5. Conclusion

This paper discusses the combination of a variety of sentiment analysis methods and related models to explore how to make social network texts in different situations more accurate and detailed in sentiment analysis, such as BERT model or multi-modal fusion. Fine-grained sentiment analysis can accurately capture complex emotions. It can not only judge the direction of emotions, but also energize the intensity of emotions and disassemble each aspect of emotions in the text. In fact, there are still many cutting-edge technologies for fine-grained sentiment analysis, such as adversarial training and robustness enhancement. Robustness refers to the ability of the model to maintain stability and accuracy in the face of noise, interference, abnormal input or adversarial attacks. When placed in a text processing task, there may be typos, network language, advertising interference, etc. In the text, which may affect the model's judgment of the text. Adding enhanced robustness can improve the model's ability to deal with such problems, and still correctly extract and judge emotions. Adversarial training can improve the robustness of the model by actively generating adversarial samples. Of course, there are other techniques to enhance robustness, such as self-supervised learning and domain adaptive training. This paper does not further explore these issues. The sentiment analysis processing method in social network text is becoming more and more comprehensive, specific and efficient, but how to protect the privacy of users while analyzing is a new challenge. The relevant privacy protection mechanisms include federal learning and differential privacy. The combination of the two can form a double privacy barrier, but there are still noise interference and efficiency problems, which need further research and analysis.

References

[1] Li, X., Bing, L., Zhang, W., & Lam, W. (2019). Exploiting BERT for end-to-end aspect-based sentiment analysis. *arXiv preprint arXiv:1910.00883.*

[2] Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. *European Language Resources Association (ELRA), 6*, 417-422.

[3] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070.*

[4] Doris, L., & Broklyn, P. (2024). Sentiment Analysis for Market Forecasting Using Machine Learning. *Machine Learning.* Advanced publication online. https://www.researchgate.net/publication/385438082_SENTIMENT_ANALYSIS_FOR_MARKET_FORECASTING_USING_MACHINE_LEARNING

[5] Chen, Y. (2015). *Convolutional neural network for sentence classification* [Master's thesis, University of Waterloo].

[6] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In K. Knight, A. Nenkova, & O. Rambow (Eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1480–1489). Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-1174

[7] Wang, C., Jiang, F., & Yang, H. (2017). A hybrid framework for text modeling with convolutional RNN. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2061–2069). Association for Computing Machinery. https://doi.org/10.1145/3097983.3098140

[8] Zhou, Y., Xu, J., Cao, J., Xu, B., & Li, C. (2017). Hybrid attention networks for Chinese short text classification. *Computación y Sistemas, 21*(4), 759-769.

[9] Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (2019). What does bert look at? an analysis of bert's attention. *arXiv preprint arXiv:1906.04341.*

[10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*, 6000-6010.

[11] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 4171-4186.

[12] Lample, G., & Conneau, A. (2019). Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291.* https://doi.org/10.48550/arXiv.1901.07291

[13] Ren, J. (2024). Multimodal Sentiment Analysis Based on BERT and ResNet. *arXiv preprint arXiv:2412.03625.* https://doi.org/10.48550/arXiv.2412.03625