Aspect Term Extraction via the Fusion of Domain-Specific and Implicit Aspect Information

Xiao Wang and Wen Rui Lu

January 6, 2024

Aspect Term Extraction via the Fusion of Domain-specific and Implicit Aspect Information

Xiao Wang ,Wen Rui Lu

Shanxi Information Industry Technology Research Institute Co., Ltd, Taiyuan 030012, China

Correspondence should be addressed to Xiao Wang:A127127000@163.com

Abstract:Aspect term extraction is primarily focused on extracting the main topics described in a text, and it has been widely applied in areas such as e-commerce and sentiment analysis. Most existing methods mainly focus on extracting features from one aspect of ambiguous information or professional domain information, which can result in loss of information from the input sentence sequence. Therefore, in this paper, we propose the Fusion of Domain-specific and Implicit Aspect Information (FDIA) model to improve the accuracy of aspect term prediction. In the FDIA model, firstly, in the representation layer, LDA and VAE-encoder are used to capture the ambiguous feature information of the input sentence sequence and concatenate them to obtain the ambiguous representation of the input sequence. At the same time, professional domain corpus is utilized to obtain the professional information representation of the input sequence. Next, attention mechanism is used to calculate the ambiguous and professional domain feature representations of the input sentence sequence. Finally, through deep mining by BILSTM and Convolution, the output label sequence is obtained.

Keywords:Aspect Term Extraction FDIA Model LDA VAE-encoder BILSTM Convolution Neural Network

1 Introduction

With the development of high technology, Natural Language Processing(NLP) technology has been widely applied. Aspect-based sentiment analysis, as a highly concerned field, plays a crucial role in the development of sentiment analysis. Sentiment word extraction, as a subtask of aspect-based sentiment analysis, is mainly used to extract topic words related to language descriptions in sentences [1-4]. Aspect words in sentences include vague words, clear words, domain words, and general words on one hand, and obscure words on the other. Clear words can be clearly understood in sentences, while vague words are more implicit [5-6].For example, "my fan has good heat dissipation effect", where "fan" is a clear aspect word; "my computer is expensive", where " price" is an aspect word but vague. General words appear in our daily life in familiar forms, while domain words appear in sentences as professional terms with industry backgrounds [7].For example, "I am a fan of this celebrity and I really like him", where "fan" is not an aspect word as a general word; "The fan of this device is of poor quality", where "fan" is an aspect word as a proper noun.

In the past, much of the work on aspect term extraction focused on designing rules for feature extraction, but this type of method is

Tabel 1

Aspect term extraction	Domain embeddings	Implicit aspect
A:The fan of this device is relatively expensive.	fan	price
B:I have just optimized of the Lion , and this feature just meets your needs.	Lion	performance
C： The case of medical devices is very strict.	case	dimensions

costly [8-9]. For a long time, people were enthusiastic about using statistical machine learning to extract aspect terms [10]. Statistical machine learning especially requires relatively accurate manual features as support, with strict conditions for feature extraction [11]. With the widespread application of deep learning technology in various fields, aspect term extraction has attracted the research interest of experts and scholars. In terms of deep learning technology, experts and scholars have conducted research and applications from different perspectives in aspect term extraction [12]. Given a learning sample, deep learning technology can mine comprehensive features, so that the final model can efficiently complete the extraction task. In aspect term mining, the ability of the model to capture the deep meaning of sentences is crucial. In addition, when extracting aspect terms involving domain words in contexts containing vague words, there is often insufficient ability. Although current aspect term extraction technology has made great progress, there are still some shortcomings:

In the context of domain words, vague words often exist. For example, in sentence A in Table 1, in the field of computer science, "fan" is a professional domain word, while "price" exists as an implicit word. In order to improve prediction accuracy, it is necessary to establish the association relationship between domain words and vague words. However, current work mainly focuses on mining domain words or vague words [13-16], and the effect of mining the association relationship between domain words and vague words is poor. In most works, when inputting representations for domain words and vague words, they are often separated, so the results of their training do not have a semantically fused representation of domain words and vague words.

If the relationship between domain words and vague words can be mined, it can greatly improve the recognition accuracy of aspect terms. For example, in sentence B in Table 1, when the feature of the vague word "performance" is ignored, the domain word "Lion" may be understood as an animal, such as an elephant or a tiger. Therefore, in the field of aspect term extraction, previous methods that separately studied the extraction of domain words and vague words may incorrectly predict aspect terms in sentence B, resulting in failing to recognize "Lion" as an aspect term here. Similarly, in sentence C in Table 1, if only the domain word "case" is mined for aspect terms, the final prediction result will lose relevant features about "dimensions". Therefore, previous research that did not consider vague word features would all lead to incorrect prediction results for related aspect terms in sentences A, B, and C in Table 1.

In summary, the integration of domain-specific words and vague words plays a key role in capturing the meaning of sentences. In [17], Xu et al. only use the features of professional words and general words as key information for sentence representation to obtain sentence information. Using professional words or vague words unilaterally as comprehensive information for sentences will lead to deviations in final prediction results.

In this paper, we propose the FDIA model to improve the accuracy of aspect term extraction. We predict aspect terms by considering comprehensive features that fuse professional domain words and vague word information, making the word representation more comprehensive. The overall structure of the model in this paper includes five parts: Input Layer, Embedding Layer, BILSTM Layer, Convolution Layer, and Output Layer. The input sentence is first processed by a feature extractor, which includes LDA fuzzy feature extraction [18-19], domain word representation [20], BERT representation [21], VAE-encoder [22], and ATTENTION layer [23-25]. Then it is processed by BILSTM for long-distance feature extraction [26-28], and finally obtains a feature vector of comprehensive information through the Convolution layer [29-31]. In the feature extractor, LDA helps to extract fuzzy information in the sentence, and VAE-encoder is helpful in obtaining hidden information. By splicing these two types of information, we can obtain a fuzzy word comprehensive feature with contextual information. Under the attention mechanism, the fuzzy word comprehensive feature and the domain word will obtain the feature related to the association between domain words and vague words. In Section 3.1, we briefly describe the problem, and in Section 3.2, we provide a detailed description of the model structure.

The main contributions of this paper can be summarized as follows:

In previous aspect term extraction work, only vague information or domain information was focused on when obtaining aspect terms, resulting in significant information loss during the extraction process. By fusing vague word information and domain word information and using attention mechanism to enable aspect terms to focus on the comprehensive information of domain words and implicit words, the acquisition of aspect terms becomes more accurate.

LDA is used to extract vague word information and the representation of vague information processed by the variational autoencoder is spliced to obtain vague word feature information about context. Then, the spliced vague word feature information is used for attention calculation with sentences containing domain word feature information to obtain sentence features with stronger comprehensive information, enabling the mining of deeper level information. The remaining structure of this paper is as follows: The related work section includes some progress in aspect term extraction; The model description section includes the model structure; The conclusion section summarizes the article.

2 Related Work

With the given set of statements, aspect term extractionWith the given set of statements, aspect term extraction technology can achieve the function of key information extraction. In most previous works, aspect terms in the text were simply extracted without considering the coexistence of vague word information and domain word information [32]. However, as model understanding ability continues to improve, it is no longer as difficult to consider both vague word information and domain word information simultaneously. In particular, the rapid development of pre-trained models has made it more convenient to solve the problem of fusing vague word and domain word information, and has also reduced the bias of model predictions.

In this paper, we focus on aspect term extraction. In many research directions in NLP, the initial methods were almost all implemented based on manually extracted features [33-34]. With the development of computer technology, neural network-based methods have been favored by experts and scholars. On the premise of above feature extraction, the final category is predicted through conditional random fields, maximum entropy, etc.

However, in previous work, only domain word information or vague word information was considered for aspect term extraction. In [14,17], only domain words or common words were considered unilaterally without synthesizing vague word information, resulting in the model only learning unilateral features of domain words or common words during aspect term extraction, thus reducing the prediction accuracy of the model. In [35], only explicit words or implicit words were considered, with more focus on capturing the characteristics of implicit information and only considering the mining of implicit words. In [36], knowledge graph related to domain words was introduced to enhance relevant feature information, but only the interdependence relationship within domain words was considered, while the relationship between domain words and implicit words was ignored. In [37], the author Xu, Qiannan et al. introduced viewpoint information to enhance the mining of implicit words in order to strengthen the feature information of implicit words, but the mining of vague word feature information was still insufficient.

Based on the above research shortcomings, the FDIA model is based on the above research shortcomings, the FDIA model is proposed in this paper to solve these existing problems. Firstly, by mining the features of implicit words and domain words, then calculating the relationship between implicit word information and domain word information to enhance aspect term information and complete accurate mining of aspect terms.

3 Model Description

3.1 Problem Description

Suppose there is a language sequence consisting of n words $X = \{x_1, x_2, \cdots x_{n-1}, x_n\}$,we aim to obtain another language sequence through sequence labeling task $Y = \{y_1, y_2, \cdots y_{n-1}, y\}$.Here, $y_n \in \{B, I, O\}$,B represents the beginning identifier of the extracted aspect term, We represents the internal word identifier of the extracted aspect term, and O represents the identifier that does not belong to the aspect term. For example, in Figure 1, after inputting the sentence sequence into the model for prediction, the corresponding labels for the words in the input sequence are obtained. The label of the aspect term "fan" is "fan", while the labels for other words that do not exist as aspect terms are "O".

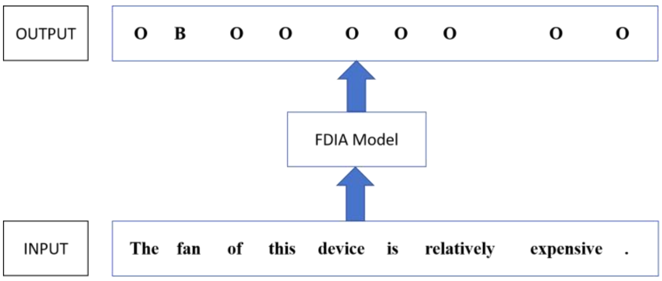

Fig.1. Sequence Labeling Diagram

3.2 Model

We propose the FIDA model to predict aspect terms. The overall structure of the model is shown in Figure 2. The FIDA model includes an input layer, a representation layer, a BILTM layer, a Convolution layer, and an output layer. The input layer is the sentence sequence. The representation layer separately obtains the implicit representation of the input sequence through domain word embedding v_n^d, LDA-extracted related words about implicit words, and BERT representation v_n^L. It also obtains the implicit representation v_n^V of the input sequence directly through VAE-encoder. Then, it splices v_n^L and v_n^V together to obtain the comprehensive implicit information representation v_n^{LV} of the input sequence. After that, the attention mechanism is used to calculate the association between the domain information representation and the implicit information representation of the input sequence to obtain the comprehensive representation v_n^{dLV} of the input sequence. The detailed structure of the representation layer is shown in Figure 3. Finally, the comprehensive representation v_n^{dLV} is further mined for deep information features of sentences through the BILTM layer and Convolution layer to make predictions more accurate.

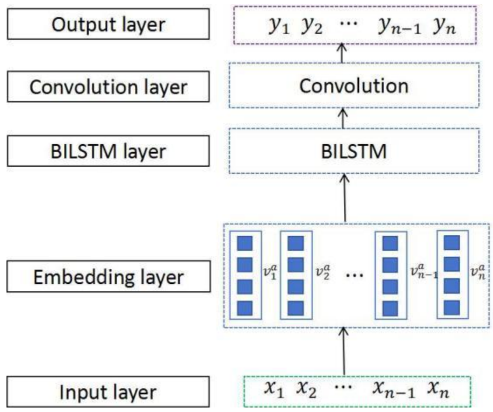

Fig.2. Model Structure Diagram

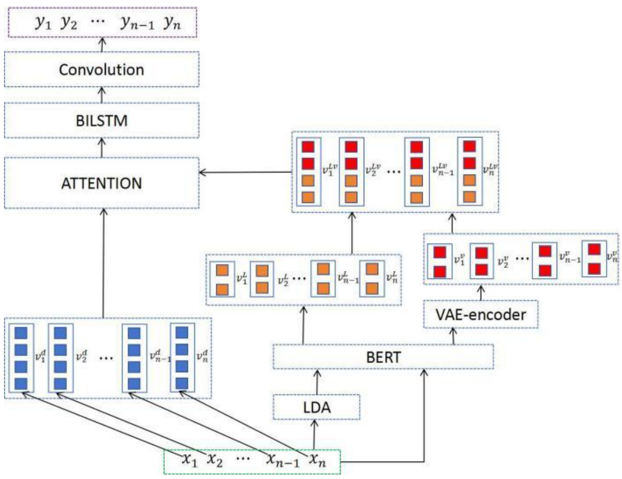

Fig.3. Detailed Structure Diagram

3.2.1 Representation Layer

The input sentence sequence is represented by professional domain words, and the corpus of professional domain word representation can be determined based on the extracted aspect term domain. Then, embedding representation can be obtained for the professional domain. The fuzzy information features v_n^{LV} obtained through LDA, BERT, and VAE-encoder operations are represented by attention mechanisms as shown in equations (1)-(4):

$$v^{LV} = v^L \oplus v^V \tag{1}$$

$$\alpha_n^{LV} = \emptyset(a v_n^{LV} + b) \tag{2}$$

$$\alpha_n^d = \emptyset(a v_n^d + b) \tag{3}$$

$$`v_n^a = \alpha_n^{LV} v_n^{LV} + \alpha_n^d v_n^d \tag{4}$$

In equation (1), represents the feature representation of the input sentence sequence extracted by LDA, v^V represents the feature representation of the input sentence sequence extracted by VAE-encoder, and $\oplus$ represents the splicing operation. In equations (2) and (3), b is obtained through training, and $\emptyset$ is a nonlinear function. In equation (4), v_n^a represents the nth element of the comprehensive representation of the input sentence sequence obtained by the representation layer.

3.2.2 BILSTM Layer

Since convolutional neural networks have a powerful feature extraction ability, the output of the BILSTM layer is used as the input to the convolutional neural network. After passing through the first layer of the convolutional neural network, the output is represented by the following equation:

$$\mathrm{First\_Layer(i,j)} = \mathrm{RELU}\left(\sum_{n=1,m=1}^{n,m} v_n^{Ba}(n,m) \cdot \mathrm{weight}(i-n+1, j-m+1)\right) \quad (11)$$

Where $\mathrm{First\_Layer(i,j)}$ represents the element at the i-th row and j-th column after the output of the first layer of the convolutional layer, $\mathrm{weight}(i-n+1, j-m+1)$ denotes the element at the i-n+1-th row and j-m+1-th column of the convolutional kernel, and RELU stands for the Rectified Linear Unit activation function.

The second layer of the convolutional layer is represented by formula (11), and similarly for the third layer. The third layer of the convolutional layer is a fully connected layer, as shown in formula (12):

$$v^c = RELU(W * \mathrm{Second\_Layer(i,j)} + \mathrm{b}) \quad (12)$$

Where W is a trainable weight matrix, b is a trainable bias parameter, and v^c represents the output label information representation of the input sentence sequence.

3.2 Training

The model's objective function is represented by formula (13), and the optimal parameters of the model are obtained using gradient descent.

$$\mathrm{Cross\_Loss} = -\sum_{s=1}^{S}\sum_{w=1}^{W} p(y_{s,w}|x_w)\log(p(y_{s,w}|x_w)) - \frac{||\mathrm{weight}||^2}{\mathrm{WS}} \quad (13)$$

Where S represents the total number of sentences in the training sample, W represents the total number of words in the training sample, Cross_Loss represents the optimization value of the objective function, $\frac{||\mathrm{weight}||^2}{\mathrm{WS}}$ represents the penalty term, and weight represents the trainable weight matrix.

4 Conclusion

In this paper, we first analyzed multiple articles from different perspectives. Then, we enhanced aspect word feature information by considering the correlation between fuzzy word information and domain word information to achieve comprehensive information mining of aspect words. To address the limitations of most current works in aspect word extraction that mainly focus on extracting either fuzzy or domain information separately, we propose the FDIA model. In the FDIA model, firstly, in the representation layer, LDA and VAE-encoder are used to capture the fuzzy features of the input sentence sequence and concatenate them to obtain the fuzzy representation of the sequence; at the same time, professional domain lexicons are used to obtain the professional domain representation of the input sequence. Next, attention mechanism is used to obtain the fuzzy and professional domain feature representation of the input sentence sequence. Secondly, through deep mining by BILSTM and Convolution Layer, the output label sequence is obtained. In this work, we mainly focused on theoretical analysis, and will conduct specific experiments in future work.

REFERENCES

[1].Yadav, Rohan Kumar, et al. "Positionless aspect based sentiment analysis using attention mechanism." Knowledge-Based Systems 226 (2021): 107136.

[2].Saberi, Bilal, and Saidah Saad. "Sentiment analysis or opinion mining: A review." Int. J. Adv. Sci. Eng. Inf. Technol 7.5 (2017): 1660-1666.

[3].Chen, Hao, et al. "Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction." Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2022.

[4].Dadap, Maureen Kate, et al. "Aspect-Based Sentiment Analysis Applied in the News Domain Using Rule-Based Aspect Extraction and BiLSTM." 2023 IEEE 6th International Conference on Computer and Communication Engineering Technology (CCET). IEEE, 2023.

[5].Zainuddin, Nurulhuda, Ali Selamat, and Roliana Ibrahim. "Improving twitter aspect-based sentiment analysis using hybrid approach." Intelligent Information and Database Systems: 8th Asian Conference, ACIIDS 2016, Da Nang, Vietnam, March 14–16, 2016, Proceedings, Part I 8. Springer Berlin Heidelberg, 2016.

[6].Mankar, Sayali A., and M. Ingle. "Implicit sentiment identification using aspect based opinion mining." Int. J. Recent Innov. Trends Comput. Commun 3 (2015): 2184-2188.

[7].Zhang, Zusheng, et al. "TADC: A topic-aware dynamic convolutional neural network for aspect extraction." IEEE Transactions on Neural Networks and Learning Systems (2021).

[8].Shu, Lei, Hu Xu, and Bing Liu. "Lifelong learning CRF for supervised aspect extraction." arxiv preprint arxiv:1705.00251 (2017).

[9].Liu, Qian, et al. "A logic programming approach to aspect extraction in opinion mining." 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). Vol. 1. IEEE, 2013.

[10].Li, Fangtao, et al. "Structure-aware review mining and summarization." Proceedings of the 23rd international conference on computational linguistics (Coling 2010). 2010.

[11].Liu, Ning, and Bo Shen. "Aspect term extraction via information-augmented neural network." Complex & Intelligent Systems 9.1 (2023): 537-563.

[12].Yang, Yinfei, et al. "Aspect extraction from product reviews using category hierarchy information." Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. 2017.

[13].Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global vectors for word representation." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014.

[14].Poria, Soujanya, Erik Cambria, and Alexander Gelbukh. "Aspect extraction

for opinion miningwith a deep convolutional neural network." Knowledge-Based Systems 108 (2016): 42-49.
[15].Feng, Yuhao, et al. "Target-specified sequence labeling with multi-head self-attention for target-oriented opinion words extraction." Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2021.
[16].Maylawati, Diah Hevyka, Warih Maharani, and Ibnu Asror. "Implicit aspect extraction in product reviews using FIN algorithm." 2020 8th International conference on information and communication technology (ICoICT). IEEE, 2020.
[17].Xu, Hu, et al. "Double embeddings and CNN-based sequence labeling for aspect extraction." arxiv preprint arxiv:1805.04601 (2018).
[18].Seymore, Kristie, and Roni Rosenfeld. "Learning hidden Markov model structure for information extraction." (1999).
[19].Shams, Mohammadreza, and Ahmad Baraani-Dastjerdi. "Enriched LDA (ELDA): Combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction." Expert Systems with Applications 80 (2017): 136-146.
[20].Wang, Tao, et al. "Product aspect extraction supervised with online domain knowledge." Knowledge-Based Systems 71 (2014): 86-100.
[21].Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arxiv preprint arxiv:1810.04805 (2018).
[22].Miao, Yishu, Edward Grefenstette, and Phil Blunsom. "Discovering discrete latent topics with neural variational inference." International conference on machine learning. PMLR, 2017.
[23].Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
[24].Yang, Zichao, et al. "Hierarchical attention networks for document classification." Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016.
[25].Hao, Yanchao, et al. "An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017.
[26].X. Li and W. Lam. "Deep multi-task learning for aspect term extraction with memory interaction." Proceedings of the 2017 conference on empirical methods in natural language processing. 2017.
[27].Joshi, Vaishali M., et al. "Deep BiLSTM neural network model for emotion detection using cross-dataset approach." Biomedical Signal Processing and Control 73 (2022): 103407.
[28].Sangeetha, J., and U. Kumaran. "A hybrid optimization algorithm using BiLSTM structure for sentiment analysis." Measurement: Sensors 25 (2023): 100619.
[29].Park, Sungrae, Wonsung Lee, and Il-Chul Moon. "Efficient extraction of domain specific sentiment lexicon with

active learning." Pattern Recognition Letters 56 (2015): 38-44.
[30].Jihan, Nadheesh, Yasas Senarath, and Surangika Ranathunga. "Aspect extraction from customer reviews using convolutional neural networks." 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer). IEEE, 2018.
[31].Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." Communications of the ACM 60.6 (2017): 84-90.
[32].Yekrangi, Mehdi, and Nikola S. Nikolov. "Domain-Specific Sentiment Analysis: An Optimized Deep Learning Approach for the Financial Markets." IEEE Access (2023).
[33].Chen, Zhiyuan, Arjun Mukherjee, and Bing Liu. "Aspect extraction with automated prior knowledge learning." Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014.
[34].Chen, Zhiyuan, et al. "Exploiting domain knowledge in aspect extraction." Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013.
[35].Yu, Li, and Xuefei Bai. "Implicit aspect extraction from online clothing reviews with fine-tuning BERT algorithm." Journal of Physics: Conference Series. Vol. 1995. No. 1. IOP Publishing, 2021.
[36].Howard, Phillip, et al. "Cross-domain aspect extraction using transformers augmented with knowledge graphs." Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022.
[37].Xu, Qiannan, et al. "Non-negative matrix factorization for implicit aspect identification." Journal of Ambient Intelligence and Humanized Computing 11 (2020): 2683-2699.