Research on the Extraction Method of Explicit Attribute-View Pairs in Product Review

Yafang Jia 1,a Shaofu Lin 2,3,b
1 College of Software, Beijing University of Technology, Beijing, 100124, China
86+18811423830
2 Beijing Advanced Innovation Center for Future Internet Technology, Beijing, 100124, China
3 Beijing Institute of Smart City, Beijing University of Technology, Beijing, 100124, China
College of Software, Beijing University of Technology, Beijing, 100124, China
86+67391304

a 13243168951@163.com  blinshaofu@bjut.edu.cn

ABSTRACT: The fine-grained affective analysis of product reviews is a fine-grained mining of the content of reviews, which has important research significance. The explicit attribute-viewpoint extraction is one of its key research issues. Due to the complex structure and colloquial features of online product reviews, traditional models are not ideal for the extraction of explicit attribute-views pairs. To solve this problem, this paper proposes an extraction method based on deep learning and integrating positional relationship information. By using the Bi-directional Long Short-Term Memory Network to overcome the problem of long distance dependence, the author makes full use of the context, combines the attention mechanism to reduce the noise weight of sentence-level informal text, and integrates the feature of position relation to enrich the feature information. The experimental results show that compared with other models, the network model proposed in this paper has improved the call rate and F1 value.

1. INTRODUCTION

With the rapid development of the Internet era, online shopping has become an important means of daily consumption. More and more users tend to express their opinions on products through online comments. The comment text contains a user of a lot of information for the opinions of products. How to obtain useful information from these large-scale data information becomes an urgent problem to be solved. In this context, sentiment analysis research came into being[1]. Sentiment analysis technology, also called opinion mining, mainly uses computers to dig out people's emotions or opinions about products. In previous research work, scholars are more inclined to conduct coarse-grained sentiment analysis. They only classify text information from three dimensions: positive, neutral and negative, which can no longer meet the needs of users and businesses. As a result, more and more scholars are beginning to pay attention to fine-grained sentiment analysis. The fine-grained sentiment analysis is to classify text information at word level and sentence level to obtain valuable opinions in text content. It is more difficult than coarse-grained sentiment analysis at text level, and has important research significance[2]. The fine-grained sentiment analysis mainly includes the following three main
tasks: emotional information extraction, emotional information classification, and emotional information retrieval and induction. Information extraction is the lowest-level task of sentiment analysis, and the extraction of attribute-view words as a sub-task of information extraction is a hot topic studied by scholars in recent years. This paper mainly studies the extraction of explicit attribute-view pairs.

2. Related work

Attribute words in product reviews are also called feature words and evaluation objects. Generally, they can be divided into explicit attribute words and implicit attribute words. Explicit attribute words appear explicitly in product reviews. The main work is to extract the attribute words and the explicit word pairs that are correctly matched in the product reviews.

Hu et al. first proposed the extraction of attribute-viewpoint pairs, and considered that the nouns with higher frequency of occurrence were often candidates for evaluation objects [3]. In the process of extracting the collocation relation of attribute-view pairs, the common methods include the following: Literature [4] used statistical methods to extract attribute words in the comment text according to association rules, and then extracted corresponding viewpoint words according to the proximity principle. Literature [5] and [6] extracted the product attribute words in the comment text by artificially constructing specific product attribute sets, and then extracted the adjectives adjacent to the attribute words as the comment words. Literature [7] used the sentiment dictionary and the topic dictionary combined with the context information to extract the attribute-viewpoint binary group. The advantage of this method was that the extracted two-tuple can directly capture the relationship between the two and improve the relationship, thus improving the performance of relational extraction. Literature [8] used point mutual information to extract attribute-viewpoint pairs. In addition, the method clustered its attributes according to sentiment words, and defines a label for each type of attribute for extracting omitted attribute words. The above methods belong to the category of unsupervised learning, and do not need to mark the corpus. The core idea is that there is no need to define the type system of extracted object relationship in advance, which is applicable to a wide range of fields. Most of these methods use optimized clustering algorithms, but the threshold setting of clustering algorithms has always been a difficult place for this method. Moreover, unsupervised learning methods lack objective evaluation criteria, and the results obtained often cannot meet the actual needs. In supervised learning, scholars mainly used machine learning to learn the marked text corpus, and then extracted the object relationship. Compared with unsupervised learning, it started relatively late, but the extraction performance of object relationships has been greatly improved, its core problem is the selection of effective features. Literature [9] proposed a set of supervised learning algorithm model for the extraction of attribute-viewpoint pairs. The idea of this algorithm was to learn candidate nodes of attribute words in tagged corpus, which was far superior to unsupervised learning in extracting attribute-view pairs. In literature [10], the problem of extracting attribute words and opinion words was modeled as a sequence labeling problem, and the conditional random field model (CRF) method was introduced. Compared with the method in literature [9], the extraction performance has been greatly improved. Moreover, this method was also applicable to relation extraction in other fields. The literature [11] combined the CRF with the rules to extract the attribute words in the comment text, and used the proximity principle and the dependency relationship to extract the corresponding viewpoint words. Literature [12] proposed a method of extracting attribute words of products by using hidden Markov model, and describing some dependence of attribute words and opinion words, and then extracting attribute-view pairs based on the consistency of context information. At present, the method based on in-depth learning is relatively less applied in the field of explicit attribute-views extraction. Therefore, this paper proposes a method based on in-depth learning and fusion of positional relationship information, which has been verified in real data. The model is described in detail in the following article.
3. Recognition Method of Attribute Words and View Words Based on CRF and Bi-LSTM

For the extraction of explicit attribute view pairs, the identification of attribute words and opinion words is a very important and indispensable step, which will affect the extraction result of explicit attribute view pairs. As the research on this aspect is relatively mature, this paper does not focus on this research. Bi-directional Long Short-Term Memory (BI-LSTM) combined with Conditional Random Fields (CRF) model is used to identify attribute words and opinion words. Its specific model architecture diagram is shown in Figure 1:

![BI-LSTM and CRF Model architecture](image)

The input layer is entered with word features and their corresponding labels. For example, “the fingerprint of the mobile phone is unlocked very quickly”, and the text is first marked in the IOB three-label format on the input layer to convert it to “hand O, machine O, referring to B_P, pattern I_P, solution I_P, lock I_P, very-B_A, fast-I_A”. Then character features are converted into corresponding randomly generated word vectors, and word labeling is converted to corresponding ID input to BI-LSTM layer. The BI-LSTM layer automatically extracts features based on contextual information to get a better representation of the original text. Then, the annotation problem is converted into a multi-classification model by using a full connection layer to get the class output of the same dimension as the number of tagged labels. The label values obtained here are only local optimum labels, without considering the global optimum and the collocation between labels, such as the case that B can not be followed by B after labeling. Therefore, the output of the sequence can be maximized by the CRF layer. Finally, the final sequence annotation results can be obtained by the Viterbi algorithm in the decoding layer, and the attribute words and opinion words in the comment text can be obtained by processing the results.

4. Extraction of Explicit Attribute-View Pairs Based on in-depth Learning

For the extraction of explicit attribute-view pairs, this paper proposes a Bi-LSTM combined attention mechanism and a network model of position prior feature information (hereinafter referred to as BLAP network) based on the LSTM network which performs well in processing sequence information. At the same time, this paper adopts the method of simultaneous extraction of attribute words and opinion words. By constructing the BLAP network model and then preprocessing the comment text, the attribute-view pairs extraction is converted into the text category judgment problem. Details is shown in 4.3.

4.1 LSTM network

Because LSTM network has the ability of long-term memory of information, it has a unique advantage in processing sequence type data. It has been widely used in various tasks in the field of Natural Language Processing (NLP). The memory function of LSTM is due to the addition of forgetting gate, input gate and output gate to the traditional recurrent neural network. The network unit structure of LSTM is shown in Figure 2.
Figure 2: LSTM network unit structure diagram

Where $\otimes$ stands for element-based multiplication, $\oplus$ stands for element-based addition, $\sigma$ stands for sigmoid activation, and $\tanh$ is the activation function. For the current time $t$, the update inside the LSTM unit can be expressed by the following formulas:

$$f_t = \sigma(W_f [h_{t-1}, x_t] + b_f)$$  \hspace{1cm} (1)
$$i_t = \sigma(W_i [h_{t-1}, x_t] + b_i)$$  \hspace{1cm} (2)
$$\tilde{C}_t = \tanh(W_c [h_{t-1}, x_t] + b_c)$$  \hspace{1cm} (3)
$$C_t = f_t \ast C_{t-1} + i_t \ast \tilde{C}_t$$  \hspace{1cm} (4)
$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$$  \hspace{1cm} (5)
$$h_t = o_t \ast \tanh(C_t)$$  \hspace{1cm} (6)

Where $x_t$ represents the input at time $t$, $h_{t-1}$ represents the output at time $t-1$, $w_f$, $w_i$, $w_c$, and $w_o$ are the parameter matrix, and $T$ represents the offset value. Figure 2 shows the output of $f_t$, forgetting the door, which is the information that needs to be forgotten by the forgotten gate. It is the output of the input gate. $i_t$ is the part of the input gate that needs to be input, and $\tilde{C}_t$ is the new cell state generated by the input gate. After the output $f_t$ of the forgetting gate is obtained, it will be further multiplied by the output $h_{t-1}$ of $t-1$ time to complete the forgetting of information. After the output $i_t$ of the input gate is obtained, it will be further multiplied by the new cell state $\tilde{C}_t$, and then the final input information will be determined. Then, it will be added with the forgetting information to complete the updating of the cell to obtain $C_t$. Finally, the updated cell state $C_t$ is multiplied by the output information $O_t$ determined by the output gate to obtain the output $h_t$ at time $t$. Although LSTM has the function of memory, it can only remember and obtain the information of each time before the current moment and after the current moment when updating cells. In order to make full use of the information of the context, this paper chooses Bi-LSTM.

4.2 Attention mechanism

Although the Bi-LSTM network makes full use of the context information in the feature representation of comment text, there will be the problem of equal weight assignment of each feature in the final category judgment based on the feature input. For the final explicit attribute-viewpoint pair extraction, it is obvious that different features play different roles in judging the final result. For example, the location information added in the subsequent BLAP network should obviously be given higher weight. In order to be able to assign different weights to different features, the Attention mechanism is applied in this paper [13]. The Attention mechanism was first applied to the field of image processing. It was first applied by Dzmitry et al. to the machine translation task in the NLP field [14], and achieved good results. The Attention mechanism has been widely used in the NLP field. At present, the mainstream...
Attention mechanisms include soft-Attention, hard-Attention, and self-Attention. Due to its simple structure and convenient implementation, soft-Attention has been widely used. Taking Attention's original application machine translation as an example, the Attention calculation process can be abstracted into these three phases as shown in Figure 3:

Relative to the Key and Value in this article are the output of the BI-LSTM network, Query is a randomly generated matrix. At the same time, for the Attention used in this paper, the calculation of phase 1 can be further divided into two steps. Step 1 is to calculate the similarity between key and Query by means of inner product, and step 2 is to calculate weight for each word according to similarity. Phase 2 is the normalization process of weights by the Softmax function, and Phase 3 is the weighted summation of Values using Phase 2 values to obtain the final attention value. All calculations can be expressed by the following formula.

\[
    u_t = \tanh(W_t h_t + b_t) \quad (7)
\]

\[
    a_t = \frac{\exp(u_t^T w)}{\sum_t \exp(u_t^T w)} \quad (8)
\]

\[
    s = \sum_t a_t h_t \quad (9)
\]

Where \(w_t\) and \(u_w\) are weight matrices, \(b_t\) is the offset value, \(h_t\) is the output of BI-LSTM, and equation (8) is used for normalization as a whole.

4.3 BLAP Network Model

The BLAP network is a network model proposed by Bi-LSTM combined with Attention mechanism and adding position prior feature information for the extraction of explicit attribute-view pairs. The network architecture is shown in Figure 4:
The input layer consists of two parts, one is text words, the other is the distance between each word and the attribute words and viewpoint words in the current comment text. The comment text data crawled from the network cannot be directly applied to the input layer, but it needs to require some processing work. First, the word segmentation is performed with foolnltk, and then the attribute words and the opinion words other than the attribute-viewpoints required for the current judgment are respectively replaced with _pro and _att. For example, "the appearance of the mobile phone is beautiful, but the battery is not durable". In judging the appearance - the explicit attribute of beauty - the processed text becomes "the appearance of the mobile phone is beautiful, but _pro _att". In this way, the explicit attribute-view pair extraction is converted into a text classification problem, and it is judged whether the attribute word and the opinion word in each text are the correct combination. The \( W \) in the input layer represents the word, the \( P_b \) and \( P_o \) respectively represent the distance between the word distance attribute word and the opinion word. The WEmbedding of the mapping layer is used to map the words to the word vector through the Word2vec\[15\] model, and the PEmbedding is used to complete the mapping of the position information to the position vector. The mapping layer generates \( R^{row \times dim} \) vectors, \( row \) is the maximum length of the sentence, \( dim \) represents the dimension of the word vector, and then extracts the features through Bi-LSTM network structure, then weights them through Attention layer, and finally completes text classification through full connection and Output layer.

5. Experiments and results analysis

5.1 Analysis of experiments process

The overall process of the experiment is shown in the following Figure 5:
Figure 5: Experiment flow

The first step is to crawl the comment data from the major e-commerce malls through the web crawler and divide the valid comment data. In the second step, the part of the valid comment data is selected to obtain the training set of the Bi-LSTM+CRF model by manual marking and the model is trained. The third step is to test the model with the test set and continuously optimize the model to achieve the best results. In the fourth step, the effective comment data is identified by the trained Bi-LSTM+CRF model, thereby obtaining attribute words and viewpoint words. In the fifth step, the identified attribute words and the opinion words are added to the custom dictionary to segment the valid comment data by fooknltk, and then the text is preprocessed as described in 4.3, and the text is divided into a training set and a test set. In the sixth step, the BLAP network model is trained through the training set, and the model is tested through the test set and continuously tuned. In the seventh step, the extraction of the display attribute-viewpoint pair is completed by the trained optimal BLAP network model.

5.2 Data set
Since there is no recognized public data set in this field, this paper uses mobile comment data set, which is the same as most other researchers in this field. For the recognition of attribute words and opinion words, a total of 8000 preprocessed data are used to train the model and 1000 are used to test the model. For the training of BLAP network model, this paper uses 10000 data as training set and 1000 data as test set.

5.3 Evaluation index
The results of this paper are the most commonly used accuracy (P), recall (R) and F1 score (F) in the classification model. The corresponding calculation formulas are as follows:

\[ P = \frac{TP}{TP + FP} \]  
\[ R = \frac{TP}{TP + FN} \]  
\[ F = \frac{2 \times P \times R}{P + R} \]

The FP in the above formula represents a false positive example, the FN represents a false counterexample, and the TP represents a real example.
5.4 Parameter Setting of BLAP Network Model
In the experiment of this paper, the vector dimension used to represent the distance between each word distance attribute word and viewpoint word is 5, the dimension of the word vector is 100, the size of the batch is 128, and the maximum length of the sentence is 70. The number of nodes in the hidden layer of the Bi-LSTM network is 100, the dimension of the Attention layer is 200, the loss rate is 0.5, the initial learning rate is 0.001, and the optimization algorithm selects Adam, and the total number of training rounds is 20.

5.5 Experimental results and comparison models
Table 1: The experimental results and comparison models

| Methods | P    | R    | F    |
|---------|------|------|------|
| SVM     | 0.74 | 0.78 | 0.76 |
| CNN+P   | 0.86 | 0.75 | 0.77 |
| BLAP    | 0.90 | 0.85 | 0.86 |

SVM is a traditional method for extracting explicit attribute-view pairs. CNN+P is an excellent method for extracting explicit attribute-view pairs by fusing location information with convolutional neural network. BLAP is a network model proposed in this paper. From the table, it can be seen that the method proposed in this paper has improved both the precision rate and recall rate and F1 score.

6. Conclusion
In this paper, the BLAP network model based on Bi-LSTM and Attention mechanism and location feature information is proposed to extract explicit attribute-viewpoint pairs in online product reviews. It has been verified in real data. Compared with traditional models, this model avoids manual feature extraction. At the same time, compared with traditional models and other network models, the accuracy and recall rate of this model are improved. However, the proposed model is not particularly ideal for the extraction of implicit attribute-view pairs, so it can be used as the next research work.

REFERENCES
[1] Cambria E, Schuller B, Xia Y, et al. New avenues in opinion mining and sentiment analysis[J]. IEEE Intelligent Systems, 2013, 28(2): 15-21.
[2] Liu Hongyu, Zhao Yanyan, Qin Bing, et al. Evaluation object extraction and its tendency analysis. [J]. Journal of Chinese Information Processing, 2010, 24(01): 84-88+122.
[3] Hu M, Liu B. Mining Opinion Features in Customer Reviews[C]//Proceedings of the AAAI-2004. 2004:755-760.
[4] Ee Hu M, Liu B. Mining and Summarizing Customer Reviews [C] // Proc. of the 10th ACM SIGKDD Int’l Conf. on Knowledge Discovery & Data Mining, 2004: 168 ~ 177.
[5] Liu J, et al. Opinion searching in multi-product reviews[C]//Proc. of the 6th IEEE Int’l Conf. on Computer and Information Technology, 2006: 25 ~ 30.
[6] Scaffidi C, et al. Red opal: Product-feature scoring from reviews[C]//Proc. of the 8th ACM Conf. on Electronic commerce, 2007: 182 ~ 191.
[7] Li B, Zhou L, Feng S, et al. A unified graph model for sentence-based opinion retrieval[C]// Meeting of the Association for Computational Linguistics. 2010.
[8] Popescu A, Nguyen B, Etzioni O. OPINE: Extracting Product Features and Opinions from Reviews[C]//Proceedings of HLT/EMNLP-2005. 2005: 32-33.
[9] Zhuang L, Jing F, Zhu X. Movie review mining and summarization[C]//Proceedings of the CIKM-2006. 2006: 43-50.
[10] Jakob N, Gurevych I. Extracting Opinion Targets in a Single and Cross-Domain Setting with Conditional Random Fields[1]. //Proceedings of the EMNLP-2010. 2010:1035-1045.
[11] Wang Zhong-Qing, Wang Rong-Yang, Pang Lei, et al. Technical report on Suda_SAM_OMS sentiment analysis system //The 3rd Chinese Opinion Analysis Evaluation. Jinan, China, 2011:25-32 (in Chinese)
[12] Lakkaraju H, Bhattacharyya C, Bhattacharya I, Merugu S. Exploiting coherence for the simultaneous discovery of latent facets and associated sentiments //Proceedings of the 2011 SIAM International Conference on Data Mining. Mesa, USA, 2011:498-509
[13] Xu K, Ba J, Kiros R, et al. Show, attend and tell: Neural image caption generation with visual attention[1]. //International conference on machine learning. 2015: 2048-2057.
[14] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[1]. arXiv preprint arXiv:1409.0473, 2014.
[15] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[1]. arXiv preprint arXiv:1301.3781, 2013.