Extract Product Features in Chinese Web for Opinion Mining

Lizhen Liu, Zhixin Lv, Hanshi Wang
College of Information Engineering, Capital Normal University, Beijing, China
xxgccnu@126.com

***Abstract*—In sentiment analysis of product reviews, one important problem is to extract people's opinions based on product features. Through the summary of feature-level opinions, different consumers can choose their favorite products according to the features that they care about. At the same time, manufacturers can also improve the product features based on the opinions. Different words may be used to express the same product feature. In order to form a useful summary, the feature words need to be clustered into different groups based on the similarity. By analyzing the characteristics of Chinese product reviews on the Internet, a novel method based on feature clustering algorithm is proposed to deal with the feature-level opinion mining problems. Particularly, 1) features considered in this paper include not only the explicit features but also the implicit features. 2) opinion words are divided into two categories, vague opinions and clear opinions, to deal with the task. Feature clustering depends on three aspects: the corresponding opinion words, the similarities of the features in text and the structures of the features in comment. Moreover, the context information is used to enhance the clustering in the procedure. Experimental evaluation shows the outperformance of the proposed method.**

***Index Terms*—feature-level, implicit features, opinion mining**

I. INTRODUCTION

With the rapid development of the Internet, a large amount of subjective reviews are available in online forums, blogs, and shopping websites. Some researches [1, 2, 3] primarily focus on recognizing opinionated sentences or documents apart from the text segments that show subjective information . While some researches [4, 5] primarily deal with classifying sentiment orientations expressed in text. They all deal with the opinion mining based on document-level. Document-level opinion mining can classify the overall subjectivity or sentiment orientation expressed in the review content, but fails to get the sentiment associated with individual features. In recent years, many researchers focus on finer-grained opinion mining which predicts the sentiment orientation related to different features as opposed to the document-level. The researches on feature-level opinion mining rely on identifying the feature words and the corresponding opinion words. However, Chinese reviews on the Internet lack of standardization. People describe their opinions using omission and free structure, which lead to a more complicated relationship between opinions and features. Then it is hard to use syntax analysis to extract features and opinion words. Particularly, for many cases, product feature words are implicit in review sentences. A feature that does not appear but is implied in the sentence is known as an implicit feature [10]. For example, the sentence "好贵啊，买不起" (*Too expensive to afford*), "贵"(*expensive*) implies the feature "价格"(*price*). Moreover, different words may be used to describe the same product feature. For example, the words "外观"(*facade*) and "外形"(*appearance*) express the same feature. The proposed method identifies the implicit features and groups the features with high similarity into one cluster. The summary can help people scan the product reviews more quickly.

For feature-level opinion mining, the most important task is to identify the feature words and the corresponding opinion words. Liu and Hu (2004), Popescu and Etzioni (2005), Kobayashi et al. (2007), Wong et al. (2008), Qiu et al. (2009), Liu and Zhang (2010) and Zhen et al. (2011) study this problem. However, this problem is far from being solved. The proposed method uses the opinion words to extract the corresponding features, and removes the noises by the support scores and confidence scores of opinions and the corresponding features. The detailed information will be introduced in Part Ⅲ.

Recently, few studies focus on recognizing the implicit features. For example, the work in [7]completely ignores the problem of recognizing implicit features. Hu and Liu [8] partially address the implicit feature identification problem by applying the same method used for explicit feature extraction. It is unreasonable to ignore the implicit features in product review contents, because people tend to express their opinions with simple structures and brachylogies, which lead to more implicit features in reviews. Su et al. [9] try to infer the implicit features for such single-word opinions, e.g. "重" (*heavy*), using Point-wise Mutual Information (PMI) based on semantic association analysis. Zhen Hai et al. [10] use CoAR algorithm to identify the implicit features. This algorithm clusters explicit features at first. The clustering results are used to choose the representative word of the cluster to the opinion word as its identified implicit feature. However, the proposed method uses the part-of-speech dictionary and the corresponding opinion words to identify the implicit features. Experimental results in Part Ⅳ shows that our method performs well.

doi:10.4304/jsw.8.3.627-632

II. Related Work

Hu and Liu (2004) [10] proposed a technique based on association rule mining to extract product features. Their main idea is that people often use the same word when they comment on the same product feature. So the frequent item sets of nouns in reviews are likely to be product features while the infrequent ones are less likely to be product features. But the infrequent items may also be features, which offer more information. This work only finds the features that many people focus on, which is not our aim. The infrequent features are also very important for people to make a choice. So our method uses opinion words to extract the corresponding features. The relationship between opinions and features is used to remove the noises to improve the precision.

Popescu and Etzioni (2005) [11] investigated the same problem. Their algorithm requires that the product class is known. The algorithm only reckons noun/noun phrase as the candidate features. It focuses on the English reviews. It determines whether a noun/noun phrase is a feature by computing the Point-wise Mutual Information (PMI) score between the phrase and class discriminators, e.g., “of xx”, “xx has”, “xx comes with”, etc., where xx is a product class. But it calculates the PMI by searching the Web. Querying the Web is time-consuming.

Qi et al. (2008) [12] proposed a novel mutual reinforcement approach to deal with the feature-level opinion mining problem. This approach clusters product features and opinion words simultaneously and iteratively by fusing both their content information and sentiment link information. This algorithm uses the relationship between the opinions and the features to extract opinion words and feature words. But, the relationship between opinions and features is so complicated that the errors will increase accordingly with the increase of iteration times in a certain range. Our method also uses the relationship between opinions and features. However, instead of iteration, the method only uses it to remove the noises by checking the mutual confidence scores and support scores. Empirical evaluation shows good performance.

Qiu et al. (2009) [13] proposed a novel algorithm called Double Propagation. It is a state-of-the-art unsupervised technique for solving the problem. Their primary idea is that opinion words are usually associated with features in some ways. Thus, opinion words can be recognized by identified features, and features can be identified by known opinion words. So the extracted opinion words and features are utilized to identify new opinion words and new features, which are used again to extract more opinion words and features. This propagation or bootstrapping process ends when no more opinion words or features can be found. The biggest advantage of the method is that it requires no additional resources except an initial seed opinion lexicon, which is readily available. It mainly extracts noun features, and works well for medium-size corpora. But for large corpora, this method can introduce a great deal of noises (low precision), and for small corpora, it can miss important features.

Zhang and Liu (2010) [14] improved the Double Propagation. This approach uses two patterns, part-whole and “no” patterns, to increase the recall and precision. As for the low precision problem, a feature ranking approach is present to tackle it. Ranking feature candidates based on the importance consists of two factors: feature relevance and feature frequency. This algorithm models the problem as a bipartite graph and uses the well-known web page ranking algorithm HITS to find important features and rank them high. However, the patterns in Chinese corpora are very few. It is possibly because many people use concise statement to write the reviews, some of which may contain the wrong grammar. The role of the model will be restricted in Chinese corpora.

The proposed approach takes modifiers as opinion words and uses the opinion words to extract the corresponding features. The main idea is that a modifier must be used to modify something. So a modifier corresponds to a feature. It can be the whole entity or a feature of the entity. If the method cannot find the corresponding feature, it must have an implicit feature for it. Hai et al. (2011) [10] used a two-phase co-occurrence association rule mining approach to identify implicit features. In the first phase of rule generation, for each opinion word occurring in an explicit sentence, they mine a significant set of association rules of the form [opinion-word, explicit-feature] from a co-occurrence matrix. In the second phase of rule application, they first cluster the rule consequents (explicit features) to generate more robust rules for each opinion word mentioned above. Given a new opinion word with no explicit feature, they then search a matched list of robust rules, among which the rule having the feature cluster with the highest frequency weight is fired, and they assign the representative word of the cluster as the final identified implicit feature. But they do not consider the opinion words, e.g. “很好” (*very good*), “不错” (*not bad*), “还可以” (*fairish*), which can modify all of the features. This kind of opinions cannot be used for distinguishing the features and that may lead to lower precision and recall. So the proposed method divides the opinions into two categories and deals with separately to solve this problem.

III. The Proposed Method for Feature-level Opinion Mining

Feature-level opinion mining includes three steps:

- Extract the features and the corresponding opinion words.
- Cluster the features.
- Orient the opinions of features.

This paper focuses on the first two tasks. For step three, we can use a sentiment dictionary to determine the orientation of the opinions, which is not our emphasis. The first two tasks are the foundation for feature-level opinion mining.

A. Extract Opinions and Features

The proposed method uses opinion words, which are modifiers, to extract the corresponding features. The mainly idea is that each modifier is used to modify a feature, no matter what is the whole entity or a part of the entity. The method considers not only the noun/noun phrase but also the verb/verb phrase, e.g. “运行” (*running*), as features. By studying the characteristics of Chinese comments, we prefer the left relationship (features are on the left of opinion words). In Chinese reviews, people tend to use the pattern“价格有点贵”(*the price is a little expensive*) rather than the pattern “很高的价格”(*high price*). Then we take the right relationship into consideration. If neither of them is used, there will be an implicit feature for it. We do not use syntax analysis which can also solve this problem, due to the complexity of the algorithm. For normative sentences, it is hard to find the opinions and the corresponding features using syntax analysis.

Obviously, using the opinions as the feature indicator is ambiguous. This means that it is not a hard rule. We inevitably get wrong features. So removing the noises is an important task. The relationship between the opinions and features is used to solve this problem. The main idea is the words with low frequency maybe the noises. The way we remove the noises is not filtering the low frequency groups, which consist of the features and the corresponding opinions, but mutual filter the noises. Firstly, the proposed approach selects the opinions with low confidence scores, then the method checks whether the corresponding features are with low confidence scores. If they do have low confidence score, the method removes the opinion words and recalculates the Co-occurrence matrix. We repeat this procedure is based on reversing the roles of opinions and features.

The confidence score of each term is determined by the following function:

$$con(x_i) = p(x_i)/N \qquad \{x_i \in F\} \vee \{x_i \in O\} \qquad (1)$$

Here, F represents the features, O represents the opinions. N represents the number of the features/opinion.

B. Identify Implicit Features

According to the way people think, people would like to omit something which is known to everyone in conversation. This phenomenon also appears in comments. In the corpora, some opinion words cannot match the explicit features. As we have mentioned above, there must be implicit features for the opinion words. There are two kinds of implicit features. One is the entity. e.g. in the sentence “不错，可以选择购买” (*not bad, we can choose it*), there is no matched feature for the word “不错”(not bad). But we know it modifies the entity that we are concerned about. For this kind of opinion words, we cannot identify the features without the context. Therefore we call this “vague opinions”. The other one is the feature. For example, in the sentence “便宜，买的值了”(*cheap, it is worth*), the word “便宜”(*cheap*) implies the feature“价格”(*price*). This kind of opinions implies the specific features which is context-independent. This is called “clear opinions”. This paper deals with the latter kind. We replace the former with the entity.

People tend to use different descriptions to express the same feature. For example,“价格”(*price*),“价值”(*value*). So for the implicit features, it does not matter which one is selected. This is why we do not use two-phase co-occurrence association rule mining approach, proposed in [10]. Besides, for the same feature, the orientation of opinion words can be different or even opposite. For example,“有点贵，勉强可以接受”(*a litter expensive, force to accept*), and “便宜，没有这么好的东西了”(*cheap, there is no better than it*), the former sentence expresses the negative opinion on the price, while the later is positive. The part-of-speech dictionary, which includes the synonyms and the antonyms, can be used to group the opinions. The proposed method groups the opinion words utilizing the part-of-speech dictionary. The explicit features, modified by the opinion group, are the candidate set for the implicit features. We select the representative word with the highest importance as the implicit feature. The importance function is expressed as follow:

$$imp(x_i) = weight(x_i)(sup(x_i) + con(x_i)) \qquad (2)$$

$$sup(x_i) = p(x_i)/N(X) \qquad (3)$$

$$weight(x_i) = \sum_{f_i \in F(x_i)} con(f_i) \qquad (4)$$

Here, $N(X)$ represents the number of candidate features; $F(x_i)$ represents the opinion words corresponding to the feature x_i. $con(\cdot)$ represents the confidence score which is calculated by Equation(1).

C. Cluster Features

We cluster the features with high similarity into groups to form a summary because people tend to use different words to express the same feature. And K-means is used to deal with grouping. The proposed method considers three aspects of the features:

1) the similarity of corresponding opinions

As we have mentioned above, the “clear opinions” can identify the features. So the similarity of the opinion words can be used to guide the clustering. Our method calculates the similarity of the corresponding opinion words utilizing their type and the Co-occurrence matrix. The similarity of this aspect is given in Equation (5).

$$simo(x_i, x_j) = index(y_i, y_j) \cdot dis(x_i, x_j) \qquad (5)$$

$$index(y_i, y_j) = \begin{cases} 1 & \{y_i \in co, y_j \in co\} \\ 0.5 & \{y_i \in co, y_j \in vo\} \vee \{y_i \in vo, y_j \in co\} \\ 0 & \{y_i \in vo, y_j \in vo\} \end{cases} \qquad (6)$$

$$dis(x_i, x_j) = \frac{A_i \cdot A_j}{\|A_i\| \cdot \|A_j\|} \qquad (7)$$

Where, “*co*” represents the “clear opinions” and “*vo*” represents the “vague opinions”. The $y_i(y_j)$ represents the opinion word that has the highest co-occurrence frequency with $x_i(x_j)$.The parameter $dis(x_i, x_j)$ represents the Cosine distance based on the Co-occurrence matrix A.

Fig.1 shows a small Co-occurrence matrix *A* as an example.

	高 (high	快 quick	不错 not bad	好 good	便宜 cheap)
性价比 (cost performance)	54	1	5	12	0
配置 (configuration)	13	0	16	10	0
速度 (speed)	15	34	12	11	0
价格 (value)	0	0	10	4	30
外观 (appearance)	0	0	14	12	0
价钱 (price）	0	0	13	5	24

Figure1. The Co-occurrence matrix A

2) The similarity of features in text

This aspect considers the features which include the same word, e.g. “运行速度” (*running speed*) and “速度” (*speed*) both represent the same feature “速度” (*speed*). The similarity in this aspect is calculated by Set Theory.

$$simt = 2 \cdot p(x_i \cap x_j)/p(x_i \cup x_j) \tag{8}$$

$P(x_i \cap x_j)$represents the number of the words which are contained in $x_i$and x_j. $p\left(x_i \cup x_j\right)$ represents the total number of words that $x_i$and $x_j$contain.

3) The structure of features in comment

This aspect considers two indexes. One is the type of the features. As mentioned, the noun/noun phrase and verb/verb phrase may be features. The proposed method considers five types: N (noun), NV(noun + verb), V(verb),VN(verb + noun), NN(noun + noun). The other one is the location of features. The Cosine distance is used to express the similarity.

$$sims\left(x_i, x_j\right) = \frac{B_i \cdot B_j}{\|B_i\| \cdot \|B_j\|} \cdot index\left(x_i, x_j\right) \tag{9}$$

$$Lindex(x_i, x_j) = \begin{cases} 0.5 & \{l(x_i) \neq l(x_j)\} \\ 1 & \{l(x_i) = l(x_j)\} \end{cases} \tag{10}$$

B represents a matrix about the kinds of the features. $l(x_i)$ represents the location of the features. Therefore, the similarity of features is represented as follows:

$$\begin{aligned} sim\left(x_i, x_j\right) = \alpha simo\left(x_i, x_j\right) + \beta simt\left(x_i, x_j\right) \\ + \gamma sims\left(x_i, x_j\right) \end{aligned} \tag{11}$$

Where, $\alpha + \beta + \gamma = 1$. In our experiments, their values are 0.7, 0.2, 0.1 respectively, which are adopted from repeatedly experiment.

D. Clustering Enhancement

The algorithm utilizes the constructed instance representation to conduct clustering process. Our basic idea of clustering enhancement by background knowledge comes from COP-KMeans [15]. COP-KMeans is a semi-supervised variant of K-Means. Background knowledge, provided in the form of constraints between data objects, is used to generate the partition in the clustering process. One type of constraints used in COP-KMeans is the incompatibility. Incompatibility: two data objects must not be in the same cluster.

The context-dependent information is also useful in the construction of constraints. In general, a review is a collection of related sentences. We assume that there is not the same feature appearing in one review. People tend to use simple sentences when express their opinions, which is unified with our observation. For example, for an editor review on computers, reviewers may usually present their opinions on the power of the battery in a sentence, followed by their opinions on the other feature in the next sentence and they do not repeat what they have described. That’s a common case in reviews. So our approach uses the incompatibility (some features cannot be grouped into one cluster) to enhance the clustering.

IV. Experiments

This section evaluates the effectiveness of the proposed method. We begin with an introduction of the data sets and evaluation metrics. Then experimental results are shown.

A. Data Sets

We used four diverse data sets to evaluate our technique. They were obtained from a commercial web (360buy.com) that provides opinion mining services. Table I shows the domains (based on their names), the number of reviews and the number of sentences (“Revi” means reviews, “Sent” indicates sentences identified by the punctuations).

B. Evaluation Metrics

- For the extracting and identifying of the implicit features, our method uses precision and recall as the evaluation metrics.
- For the clustering, the proposed approach uses the VI (Variation of Information) as the evaluation metrics. It is an information theoretic measure that regards the system output C and the gold standard tags T as two separate clusters, and evaluates the amount of information lost in going from C to T and the amount of information gained, i.e., the sum of the conditional entropy of each clustering conditioned on the other. More formally:

$$\begin{aligned} VI(C,T) &= H(T \mid C) + H(C \mid T) \\ &= H(C) + H(T) - 2I(C,T) \end{aligned} \tag{12}$$

$H(\cdot)$ is the entropy function and $I(\cdot)$ is the mutual information. VI and other entropy-based measures have been argued to be superior to accuracy-based measures, because they consider not only the majority tag in each cluster, but also whether the remainder of the cluster is more or less homogeneous.

C. *Experimental Results and Analyses*

We first compare our method with Double Propagation in the extraction of features and the corresponding opinions. The results are presented in Table II. "*Ours*" represents our method, and "*DP*" means Double Propagation.

From the Table II, for corpora in different domains, our method outperforms or is equal to Double Propagation on recall. On "computer2", the precision even is better. This may be because of the removing of noises, based on mutual information. However, for the small scale date sets, the recall and the precisions both are low.

For identifying the implicit features, we calculate the precision and recall based on Manual annotation results. Table III shows the results. For clustering the features, figure 2 shows the value of VI on different K for the four data sets.

In the Fig.2, the best numbers of the clustering for the four data sets are 23, 33, 25, 35 respectively, which are close to the gold standard, 29, 35, 30, 30. We also verify the effectiveness of the enhancement based on the context-dependent information. Table IV shows results. "K-Means" represent the pure K-Means algorithm and "Enhance" represent the K-Means based on the knowledge that is used in the proposed method.

In the Table IV, the K-Means based on the knowledge is much better. It shows the context-dependent information is a good indicator for the cluster of features.

TABLE I.
THE DATA SETS

Date sets	Computer1	Computer 2	Phone	Camera
Revi	500	1000	1000	1000
Sent	1459	2798	3067	2674

TABLE II.
THE RESULTS OF THE EXTRACTION

Date sets		Computer1	Computer2	Phone	Camera
Precision	***Ours***	0.57	**0.69**	**0.71**	0.64
	DP	**0.59**	0.65	**0.71**	0.65
Recall	***Ours***	**0.55**	**0.60**	**0.62**	**0.58**
	DP	**0.55**	0.58	**0.62**	0.57

TABLE III.
THE IDENTIFYING OF THE IMPLICIT FEATURES

Date sets	Computer1	Computer2	Phone	Camera
Precision	0.65	0.72	0.79	0.74
Recall	0.56	0.67	0.70	0.65

V. CONCLUSIONS

Feature extraction for entities is one of the most important tasks in opinion mining. This paper proposed a novel method to deal with this task. The novel method uses the corresponding opinion words to extract features, and filters the noises according to mutual support scores and confidence scores. It also identifies the implicit features and clusters the features based on the knowledge of the context-dependent information. Experimental evaluation shows the outperformance of the proposed method. However, this method has some shortcomings. In small scale corpora, it cannot perform well. The structure of the vague opinions dictionary and the part-of-speech dictionary increase the cost of the method. In this paper, we do not consider the case that some verbs imply opinions. Next, we will study the automatic establishment of two dictionaries, improve the precision and recall for the small scale corpus and consider the influence of verbs

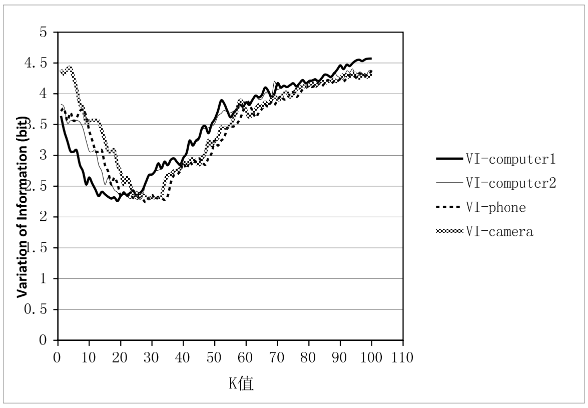

Figure 2. The results of clustering

TABLE IV
THE RESULTS OF COMPARE

Date sets		Computer1	Computer2	Phone	Camera
K		**23**	**33**	**25**	**35**
Precision	K-Means	0.54	0.64	0.63	0.72
	Enhance	**0.65**	**0.70**	**0.67**	**0.79**
Recall	K-Means	0.43	0.52	0.58	0.55
	Enhance	**0.53**	**0.62**	**0.65**	**0.64**

ACKNOWLEDGMENT

This work was supported by the Beijing Key Disciplines of Computer Application Technology, China.

REFERENCES

[1] Blair-Goldensohn, Sasha., Kerry, Hannan., Ryan, McDonald., Tyler, Neylon., George A. Reis, Jeff, Reyna., “Building Sentiment Summarizer for Local Service Reviews,” In Proceedings of the Workshop of NLPIX , WWW, 2008 .

[2] Ding, Xiaowen., Bing Liu, Philip S. Yu., “A Holistic Lexicon-Based Approach to Opinion Mining,” In Proceedings of WSDM, 2008.

[3] Chieu, Hai Leong and Hwee-Tou Ng, “Name Entity Recognition: a Maximum Entropy Approach Using Global Information,” In Proceedings of the 6th Workshop on Very Large Corpora, 2002.

[4] Hu, Mingqin and Bing Liu., “Mining and Summarizing Customer Reviews.,” In Proceedings of KDD, 2004.

[5] Kobayashi, Nozomi., Kentaro Inui and Yuji Matsumoto, “Extracting Aspect-Evaluation and Aspect-of Relations in Opinion Mining,” In Proceedings of EMNLP, 2007.

[6] Bing Liu, “Sentiment Analysis and Subjectivity,” in Handbook of Natural Language Processing, Second Edition, pp.627-665, CRC, 2010.

[7] Hu, M., Liu, B., “Mining and summarizing customer reviews,” In Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, *USA*, pp. 342–351, 2004.

[8] Hu, M., Liu, B., “Opinion feature extraction using class sequential rules,” AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs, Palo Alto, *USA* , 2006.

[9] Su, Q., Xiang, K., Wang, H., Sun, B., Yu, S., “Using pointwise mutual information to identify implicit features in customer reviews,” In Matsumoto, Y., Sproat, R.W., Wong, K.-F., Zhang, M. (eds.) ICCPOL 2006 [LNCS (LNAI), vol. 4285, pp. 22–30.Springer, Heidelberg , 2006]

[10] Zhen Hai, Kuiyu Chang, and Jung-jae Kim A. Gelbukh (Ed.), “Implicit Feature Identification via Co-occurrence Association Rule Mining,” CICLing , Part I, LNCS 6608, pp. 39 3–404, 2011.

[11] Popescu, Ana-Maria and Oren, Etzioni. 2005, “Extracting product features and opinions from reviews,” In Proceedings of EMNLP, 2005.

[12] Su, Qi., Xinying Xu., Honglei Guo, Zhili Guo, Xian Wu, Xiaoxun Zhang et al., “Hidden Sentiment Association in Chinese Web Opinion Mining,” In Proceedings of WWW , 2008.

[13] Qiu, Guang., Bing, Liu., Jiajun Bu and Chun Chen, “Expanding Domain Sentiment Lexicon through Double Propagation,” In Proceedings of IJCAI, 2009.

[14] Lei Zhang, Bing Liu, Suk Hwan, Lim Eamonn, O’Brien-Strain , “Extracting and Ranking Product Features in Opinion Documents,” Coling, Poster Volume, pages 1462–1470, Beijing, August , 2010.

[15] K. Wagsta, C. Cardie, S. Rogers, and S. Schroedl., “Constrained k-means clustering with background knowledge,” In Proceedings of the Eighteenth International Conference on Machine Learning, pp. 577–584, 2001.

Prof. Lizhen Liu holds a PhD in Computer Application from the Beijing Institute of Technology, China. She is currently a Professor at the Capital Normal University. Her research interests include text mining, sentiment analysis, knowledge acquisition, and the design of Intelligent Tutoring Systems.

She has published in journals and conferences like Knowledge and Information Systems, International Journal of Information & Computational Science, Journal of Software, IEEE World Congress on Intelligent Control and Automation, CSCWD and so on.

She served as PC of a number of international conferences, including IEEE Pervasive Computing and Application, International Symposium on IT in Medicine & Education, Computer Supported Cooperative Work in Design and so on. Prof. Lizhen, the director of intelligence science & technology department, is also a member of ACM, IEEE and CCF.

Zhixin Lv is studying for a master's degree at Capital Normal University. Now, her research interest includes Sentiment Analysis, Text Mining, and Nature Language Processing.

Dr. Hanshi Wang holds a PhD in Computer Application from the Beijing Institute of Technology, China. He is currently a lecturer at the Capital Normal University. His research interests focus on computational linguistics and natural language processing, especially unsupervised methods in the area.

He has published his important work on the famous journal of Computational Linguistics (CL), and other international conferences.