Feature-based Sentiment Analysis Approach for Product Reviews

Hanshi Wang, Lizhen Liu*, Wei Song
College of Information Engineering, Capital Normal University, Beijing, China
xxgccnu@126.com

Jingli Lu
Grasslands Research Centre, New Zealand
Janny.jingli.lu@gmail.com

***Abstract*—The researches and applications of sentiment analysis become increasingly important with the rapid growth of online reviews. But traditional sentiment analysis models have been lacking in concern on the modifying relationship between words for sentiment analysis of Chinese reviews, and limit the development of opinion mining. This paper proposes a feature-based vector model and a novel weighting algorithm for sentiment analysis of Chinese product reviews. It considers both modifying relationships between words and punctuations in review texts. Specifically, it can classify reviews into two categories, i.e., positive and negative, and can also represent the sentiment strength by adverb of degree. Moreover, a novel feature extraction method based on dependency parsing is presented to identify the corresponding aspects that opinions words modify. We conduct some experiments to evaluate our algorithms, and demonstrate that the proposed approaches are efficient and promising.**

***Index Terms*—sentiment analysis, dependency parsing, polarity classification**

I. Introduction

With the development of the Internet, it provides a new platform for people to express their attitudes, opinions and feelings. At the same time, the explosion of online reviews, presents a challenge that how people discover useful information and make use of it effectively in various situations. Sentiment analysis starts to play an important role in extract the opinions or sentiments of reviews, which is to identify and extract subjective information in text materials, such as opinions and feelings. The opinions expressed by customers can help manufactory to improve their products or services, and help potential customers make purchase decisions. Thus, accurate understanding of sentiments expressed in E-commerce websites could bring tremendous business opportunities and plays a crucial role in making effective decisions.

Recently, most current Sentiment Analysis methods have focused on analyzing English and other European languages. Very few studies have addressed sentiment analysis in Chinese language which is a morphologically rich language [1, 2]. In addition, sentiment analysis is commonly considered as a binary classification problem [3, 4]. Few researchers take the strength of sentiments into consideration. However, sentiment with the same polarity may reflect different sentiment strength that indicates the degree of positive or negative the opinion. For instance, “The screen is really cool” indicates a very strong positive opinion, whereas “The screen is good” indicates a weak positive attitude. This motivates us to conduct research in analyzing both semantic orientation and strength of Chinese Product Reviews.

In this paper, we focus on analyzing opinions expressed by customers in Chinese reviews. We propose a feature-based vector model for in-depth sentiment analysis of review texts and a novel weighting algorithm. Liu.B [5] proposes a quintuple model (object, feature, opinion, opinion holder, time) to express a direct opinion. The quintuple model mainly focuses on the polarity of the opinion on object and performs well in English. However, it can’t effectively find true feelings from Chinese reviews due to the lack of concern on modifying relation between words and complexity of Chinese. Our study not only leverages modifying relationship, but also identifies punctuation from sentiment sentence. The same sentence with different punctuation may express different emotion. For example, “He loves this dress?” is different from “He loves this dress!”. Generally, “!” express a stronger emotion, while “?” often invert the sense of the sentence. Our algorithm based on feature-based text vector model with dependency parsing to classify reviews into two categories, but also assigns sentiment strength on them.

The rest of the paper is organized as following: in the next section, related work on sentiment analysis is discussed. In Section 3, a feature-based text vector model and a novel weighting algorithm are proposed. Section 4 describes a novel feature extraction method that is based on dependency parsing. Experimental results are presented in Section 5, and we conclude in Section 6.

* Supported by the Beijing Natural Science Foundation (No. 4133084) and the Beijing Key Disciplines of Computer Application Technology.

* Corresponding Author: Lizhen Liu

doi:10.4304/jsw.9.2.274-279

II. Related Works

The approaches for sentiment classification could be roughly categorized into two groups: approaches based on machine learning techniques and approaches based on lexicon. Most studies have been focusing on training classifiers based on machine learning algorithms to classify reviews. Pang et al. [6] used three traditional machine learning methods to classify movie reviews of Internet Movie Database(IMBD), i.e., Naive Bayes, Maximum Entropy and Support Vector Machines. Their experimental results demonstrate that all these three methods outperform human-produced baselines and SVM achieves the best result. Mullen and Collier [7] employ a hybrid SVM approach by making use of potentially pertinent information, including several favorability measures of terms and knowledge of the topics. Blitzer et al. [8] investigate domain adaptation for sentiment classifiers, focusing on online reviews for different types of products. Prabowo and Thelwall [9] conducted experiments on movie reviews, product reviews and Myspace comments by combining a rule-based classifier and supervised learning algorithm, and found that the hybrid classification showed an improvement in accuracy. Machine learning approaches work well in situations where large labeled corpora are available for training and validation [10].

In contrast, lexicon-based approaches classify the sentiment of the text by analyzing the sense of the opinion words methods, which measure the polarity of text based on lexicons, such as WordNet [11], MPQA lexicon [12] and SentiWordNet [13]. Turney [14] predicates the sentiment orientation of phrases by Point-wise Mutual Information (PMI) based on pre-defined seed words, and classifies a document as positive or negative by the average semantic orientation of the phrases. Kim and Hovy [15] build three models to assign a sentiment category to a given sentence by combining the individual sentiments of sentiment-bearing words. Kennedy and Inkpen [16] classify reviews based on the number of positive and negative terms in which they contain. Devitt and Ahmad [17] explore a computable metric of positive or negative polarity in financial news text. Lexicon-based approaches outperform machine learning approaches when training corpus is not sufficiently large to accumulate the necessary feature frequency information. However, the difficulty of generating a complete sentiment dictionary hinders the application of lexicon-based approaches in sentiment analysis. Therefore, our proposed method is based on machine learning techniques.

Due to the complexity of semantic expression, simple word and phrase cannot express the semantic orientation accurately and traditional text vector model shows very poor performances in sentiment classification. In recent years, some research has been conducted to focus on the definition of opinion model. Kim and Hovy [15] describe an opinion as a quadruple (topic, holder, Claim, Sentiment). Liu [5] analyze and propose a feature-based sentiment analysis model (object, feature, opinion, opinion holder, time) to express an opinionated document. However, these definitions lack of concern in Chinese characteristics and ignore the over modifier of opinion words. For example, “The bag is a little too big”(这个书包有点太大了), the model ignore the adverbs of degree around the opinion word, cannot express the strength of sentiment accurately, and when the words like “too”, “excessively” occur around the opinion word, they may invert the opinion sense. Moreover, sentiment words can be classified into two categories, static and dynamic. Static sentiment words usually have a clear semantic orientation, such as beautiful, clever, whereas dynamic sentiment words may have opposite emotion in different environment. For example, “high quality” indicates a positive opinion, but “high price” may express a negative opinion.

In this paper, we propose a feature-based vector model for Chinese Reviews, which identifies both modifying relationships between words and punctuations. The understanding of sentence structure and content is an effective way to solve the ambiguity of emotion, which has a great impact on sentiment analysis. Furthermore, the analysis of modifying relationship contributes to identify the polarity of document and also can reflect the sentiment strength.

III. Proposed Feature-based Vector model for Chinese Sentiment Analysis

A. Feature-based Six-Tuple Vector Model

The model we propose for Chinese sentiment analysis is called feature-based six-tuple vector model. Our feature-based vector model contains both features and opinions that describe reviewers’ opinions on the features, and also contains the modifying relationship between words, and the punctuation of an opinion sentence. We define a general opinionated document D as $\{(t_1, \omega_1), (t_2, \omega_2), \ldots, (t_n, \omega_n)\}$, where ω_n is the weight of the opinion, i=1…n. t_i is a feature-based vector that represents a reviewer’s opinion on an object, as shown as follows:

$$t_i=(f, o, omc, gms, nmc, p) \tag{1}$$

In this model, f is a feature or attribute that a reviewer comments on, o is an opinion word of feature f, omc is the number of over modifier of opinion word o. Over modifier are the adverbs of over degree like “excessively”, “too”, “ultra” etc., which may invert the polarity of opinion words and have an influence on sentiment orientation of review. There are 30 frequently used over modifier in Chinese language, and Chinese like to use them to express their felling. For this reason, we extract the over modifier around the opinion word. The gms is the average score of general modifier of opinion word o, nmc is the number of negation words of opinion word o and p is the punctuation of the opinion sentence, which reflects the tone of reviewer. Generally, the number of over modifier and negation words of an opinion word is no more than 2. Therefore, we assign score (0, 1, 2) to omc and nmc. P has three different states, i.e., statement, exclamation and interrogation.

Our feature-based vector model represents the necessary information for classifying Chinese review. We define three variables to express the modifier of opinion, i.e., general adverbs of degree, over modifier and negation words. The adverbs of degree express the strength of polarity. Negation words can invert the polarity of the opinion word. Thus, accurate analysis of the modifiers of opinion words is the basis of effective sentiment classification. In this paper, we employ the Chinese sentiment lexicon generated from HowNet, which is a widely used online common-sense knowledge database unveiling inter-conceptual relations and inter-attribute relations of concepts as connoting in lexicons of the Chinese and their English equivalents. HowNet classifies adverbs of degree into six levels.

In our feature-based vector model, levels 1-5 are seen as adverbs of general degree, which can impact on the intensity of polarity, but don't invert the sense of opinion. We assign scores (1, 2, 3, 4, 5) to level 1-5, respectively. The value of *gms* is the average score of the adverbs of general degree for each opinion word. But level 6 is seen as over modifier, sometimes may invert the semantic orientation, and the number of over modifier can also influence the semantic orientation, therefore we define *omc* as the number of over modifier associated with the opinion word. Because the average score of adverbs of general degree expresses the influence on intensity of polarity better, the quantity of over modifier and negation words could indicate whether the semantic orientation is inverted. For each review, we compute the *gms* per feature. This score is used to express the intensity of polarity.

B. Feature Weighting Algorithm

TF-IDF [18] is used as a major algorithm for text classification and has gained great success. This algorithm focuses on the significant word for classification, which has a high term frequency in the target document and a low frequency in the whole collection of documents. However, sentiment is not usually highlighted by repeating the same terms. Eirinaki et al. [19] points out that noun having more adjectives to modify them are more likely to be the important and distinguishing features. Here we propose a feature weighting algorithm, called High Adverb of Degree Count (HADC) algorithm base on the feature-based vector model. For each opinion feature, HADC can be described as follows:

$$hadc_i = \frac{1 + degadv_i}{n + \sum_{i=1}^{n} degadv_i} \qquad (2)$$

where n is the number of opinion words and $degadv_i$ represents the number of adverbs of degree that are used to modify the *opinion-word*$_i$ in a text review.

In a nutshell, the main idea behind the algorithm is that sentiment words that an reviewer uses more adverbs of degree to describe is more likely to be the important element to represent the sentiment of Chinese reviews. The reason is that Chinese people usually use more adverbs of degree to emphasize emotion. For an example, 这是一台不错的电脑，我真的很喜欢它的大屏幕，色彩也非常漂亮("It is a good computer, I really love the big screen so much and color is very beautiful too") , the reviewer uses two adverbs of degree to emphasize the love for screen.

IV .SENTIMENT SIX-TUPLE EXTRACTION BASED ON DEPENDENCY PARSING

In this section, we propose an algorithm to extract sentiment six-tuple based on dependency parsing for Chinese sentiment classification. Dependency parsing has been widely used in sentiment analysis, due to its high performance of syntactic parsing at sentence level. The research result in Zhou M [20] shows that dependency parsing is more suitable for analyzing Chinese sentence than Phrase Structure Grammar. Although phrase structure grammar has achieved great success in English, it cannot effectively describe the relationship between words in Chinese sentences. The main elements of dependency structure are dependency pairs, in which one is a core word, and the other is a subordinate word. Each sentence can be represented as a parsing tree by dependent arcs that connect a core word to a subordinate word. For example, the dependency parsing tree of sentence "相机的质量非常好." (The quality of the camera is very good.) is shown in Fig.1. In dependency pair (good, quality, SBV), "good" is the core word and "quality" is the subordinate word. Dependency parsing reveals syntax structure by analyzing semantic dependency relationship between the components of language units. Therefore our extracting algorithm is based on dependency parsing.

Our sentiment six-tuple extraction algorithm has two inputs, each of which is described in the paragraphs below.

The first input is a list of sentiment words. We use four subsets of HowNet that is a Chinese Vocabulary for Sentiment Analysis, i.e., positive evaluation words subset (e.g. 漂亮(beautiful), 聪明(clever), negative evaluation words subset (e.g. 坏的(bad), 卑鄙(despicable)), positive sentiment subset (e.g. 满意(satisfied), 喜欢(like))and negative sentiment words subset(e.g. 愤怒(angry), 失望(disappointed)). These words could express strong appraisal emotions, which are the most widely used sentiment words extracted from a large number of corpus.

The second input is a list of modifier that includes negation words and adverbs of degree. Negation words can be used to invert the semantic orientation of reviews, such as "no" and "not". Adverbs of degree are words that have impact on the intensity of polarity, such as "more", "especially", "merely". HowNet sentiment word set provides a list of adverbs of degree containing 219 words. Furthermore, these words are divided into six levels according to their different intensity.

In each six-tuple, opinion feature f can represent attributes of reviewed objects, or objects themselves. Opinion features can be identified and marked by analyzing the result of dependency parsing. Considering the time complexity and computational complexity, we only focus on the dependencies that are useful for feature extraction, such as SBV, VOB, ADV, ATT, COO，CMP and HED. Firstly, if the subordinate word in SBV and VOB, and the core word in ATT is noun or pronoun, then they are more likely to be reviewed objects, because the reviewed objects are usually nouns and pronouns. Secondly, the components, which have coordinative relation (COO) with the words identified in previous step, are identified as the reviewed objects too. Finally, pronouns are replaced by the previous noun that is identified as a reviewed object.

The next step is to extract the sentiment word. We search the dependencies of the opinion feature and identify the opinion words associated with the opinion feature in the sentence. If the opinion feature is subordinate word in structure SBV, and the core word can be found in the sentiment word list, then the core word is an opinion word. Otherwise, we search for ADV, VOB and CMP relation around opinion feature and if the subordinate word in above relations can be found in sentiment word list, then the subordinate word is opinion word. If the reviewed object is subordinate word in structure SBV and the core word can be found in sentiment word list, then the subordinate word is an opinion word. If the object is the core word in structure ATT and the subordinate word can be found in sentiment word list, then the subordinate word is an opinion word. For example, "舒适的房间"(It is a comfortable room), "舒适"(comfortable) is the opinion word "房间"(room). Finally, if opinion words are in coordinative relation (COO), then the words in COO are opinion words too.

Furthermore, we identify the modifier of each opinion word by analyzing the ADV relation before an opinion word. We identify modifiers and modifying types depending on a list of modifier. The scores of *omc*, *gms*, *nmc* are initialized to zero. Each adverb and inversion word is associated with the opinion word that the modifier is the most likely to describe. For each over modifier, the score of the *omc* is increased by one. Similarly, *nmc* is increased by the number of inversion word. For each *gms* we compute the average score per general adverb of degree associated with the opinion word. Finally, we also need to identify the tone of the review sentence and assign it to p. If the punctuation is"!", *p* is set to "exclamation", else if the punctuation is"?", *p* is set to "interrogation" and *p* is set to "statement" in other cases.

V. EXPERIMENTS AND RESULTS ANALYSIS

A. Dataset and Evaluation Measures

We have conducted experiments on ChnSentiCorp to evaluate the effectiveness of the proposed method, which is a Chinese sentiment corpus provided by Songbo Tan. We have extracted 1375 documents from the corpus, which consists of customer reviews in three domains, i.e., Electric devices, E-journal, and hotel. The data in each domain contains both positive and negative documents. A brief summary of these three dataset is shown in Table Ⅰ.

TABLE I.
THE SIZE OF THREE SENTIMENT CORPORA

Date sets	Sentiment	Documents
Electric devices	Positive	356
	Negative	306
Hotel	Positive	180
	Negative	218
E-journal	Positive	139
	Negative	158

We use three evaluation measurements that are commonly used in sentiment analysis to evaluate the effectiveness of our proposed method, i.e., Precision, Recall and Accuracy. These evaluation measurements for each corpus are calculated using the following formulas:

$$Precision = \frac{number\ of\ correct\ \ predictions}{number\ of\ \ predictions} \quad (3)$$

$$Recall = \frac{number\ of\ correct\ \ predictions}{number\ of\ \ examples} \quad (4)$$

$$Accuracy = \frac{number\ of\ correct\ predictions}{number\ of\ \ examples} \quad (5)$$

B. Comparison and Analysis

In this section we combine feature-based vector model with three weighting algorithms respectively, i.e., our proposed High Adverb of Degree Count (HADC), the well-known TF and TF-IDF. We compare the performance of these three cases. As the popularity and high performance of support vector machines (SVM) for sentiment classification, we use SVM as the classifier in our experiments. We perform this by comparing the Precision and Recall in classifying the reviews as positive or negative.

Table Ⅱ presents the performance of feature-based vector model combined with three weighting algorithms in three dataset. It is shown that our proposed approach consistently yielded better results than the TF or TF−IDF algorithms. The algorithm's performance in three datasets verifies our initial intuition that opinion words which reviewers use more adverbs of degree to describe is most important element to represent the sentiment of Chinese reviews.

We have also computed the accuracy of the proposed approach in classifying the reviews as positive or negative for different data sets. Table Ⅲ shows the accuracy of the classified reviews for the various datasets with different weighting algorithms. It is clear that the HADC algorithm, which weight each six-tuple feature based on the count of the number of adverbs of degree, outperforms TF or TF−IDF algorithms. The average accuracy of HADV is 0.8997, which is three percents greater than TF (0.8654), and two percents greater than TF-IDF (0.8761).

The experimental results show that our proposed six-tuple model and HADC weighting algorithm succeeds to classify more than 88% of the reviews correctly in the worst case, and has close to 92% accuracy in the best case.

TABLE II.
THE PERFORMANCE OF FEATURE-BASED VECTOR MODEL COMBINED WITH THREE WEIGHTING ALGORITHMS IN SENTIMENT CLASSIFICATION

Date sets	Method	Positive		Negative	
		Precision	Recall	Precision	Recall
Electric devices	TF	0.8270	0.9121	0.8801	0.8069
	TF-IDF	08630	0.8196	0.8149	0.8459
	HADC	0.9156	0.8951	0.8955	0.8656
Hotel	TF	0.8556	0.8829	0.9026	0.8793
	TF-IDF	0.8913	0.8723	0.8983	0.9137
	HADC	0.9456	0.9255	0.9406	0.9568
E-journal	TF	0.8048	0.9428	0.9354	0.7837
	TF-IDF	0.8732	0.8857	0.8904	0.8783
	HADC	0.8648	0.9142	0.9142	0.8648

TABLE III.
ACCURACY OF THREE WEIGHTING METHODS FOR SENTIMENT CLASSIFICATION

	TF	TF-IDF	**HADC**
Electric devices	0.8542	0.8514	0.8914
Hotel	0.8809	0.8952	0.9190
E-journal	0.8611	0.8819	0.8888

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a feature-based vector model and a new weighting algorithm for sentiment analysis of Chinese reviews. Furthermore, an effective feature extraction method based on dependency parsing is proposed. Our proposed model considers both modifying relationships between words and punctuations in review texts. Our model can classify reviews into two categories, i.e., positive and negative, and can also represent the sentiment strength by adverb of degree. An experimental evaluation on three review data sets in different domains has shown that our algorithm achieves a high level of accuracy.

With the consideration of field dependence, our plan for future work is to investigate how to transfer sentiment classifier from one domain to another effectively.

ACKNOWLEDGMENT

This work was supported by the Beijing Natural Science Foundation (No. 4133084) and the Beijing Key Disciplines of Computer Application Technology.

REFERENCES

[1] Tan S, Zhang J (2008) An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications* 34(4):2622-2629.
[2] Wan X (2011) Bilingual co-training for sentiment classification of Chinese product reviews. *Computational Linguistics* 37 (3):587-616.
[3] Li S, Zhou G, Wang Z, Lee SYM, Wang R (2011) Imbalanced sentiment classification. In: *Proceedings of CIKM-2011*, pp 2469–2472.
[4] Tang H, Tan S, Cheng X (2009). A survey on sentiment detection of reviews. *Expert Systems with Application* 36(7):10760–10773.
[5] Liu B (2010) Sentiment Analysis and Subjectivity. In: *Indurkhya N and Damerau FJ (ed) Handbook of Natural Language Processing*, 2nd edn., pp 627–667.
[6] Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pp 79-86.
[7] Mullen T, Collier N (2004) Sentiment analysis using support vector machines with diverse information sources. In: *Proceedings of EMNLP-2004*, pp 412–418.
[8] Blitzer J, Dredze M, Pereira F (2007) Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics,* pp 440–447.
[9] Weifeng Pan, Shan Li, "Tag Ontology Automatic Building for Semantic Searching of Services: a Case Study on Mashup Services," *Journal of Computers*, Vol 7, No 12 (2012), pp. 2979-2986, Dec 2012.
[10] Prabowo R, Thelwall M (2009) Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2): 143–157.
[11] Andreevskaia A, Bergler S (2008) When specialists and generalists work together: Overcoming domain dependence in sentiment tagging. *In: Proceedings of ACL-08*: HLT, pp 290-298.
[12] Miller G, Beckwith R, Fellbaum C, Gross D, Miller K. (1990). Introduction to wordnet: An on-line lexical database. *International Journal of Lexicography*, 3(4): 235–312.
[13] Jike Ge, Zushu Li, Taifu Li, "A Novel Chinese Domain Ontology Construction Method for Petroleum Exploration Information," *Journal of Computers*, Vol 7, No 6 (2012), pp. 1445-1452, Jun 2012.
[14] Wilson T, Wiebe J, Hoffmann P (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing* (HLT/EMNLP), pp 347–354.
[15] Esuli A, Sebastiani F (2006) Sentiwordnet: A publicly available lexical resource for opinion mining. In: *Proceedings of the 5th conference on language resources and evaluation (LREC)*, Genova, Italy.
[16] Turney PD (2002) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *In Proceedings of ACL-02* ,pp 417–424.
[17] Kim SM, Hovy E (2004) Determining the sentiment of opinions. In: Proceedings of COLING-04, *20th International Conference on Computational Linguistics*, pp 1367-1373.
[18] Kennedy A, Inkpen D (2006) Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence* 22(2):110-125.
[19] Devitt A, Ahmad K (2007) Sentiment polarity identification in financial news: A cohesion based approach. In: Proceedings of ACL-07, pp 984-991.
[20] Salton G, Buckley C (1988) Term-weighting approaches in

automatic text retrieval. *Information Processing and Management* 24 (5):513–523.
[21] Eirinaki M, Pisal S, Singh J (2012) Feature-based opinion mining and ranking. Journal of Computer and System Sciences 78(4): 1175-1184.
[22] Zhou M (2000) A Block-based Robust Dependency Parser for Unrestricted Chinese Text.In: Proceeding of the second workshop on Chinese language processing: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics.

Author Biographies

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.

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 Computers, 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.

Dr. Wei Song holds a PhD in Computer Application from the Harbin Institute of Technology, China. He is currently a lecturer at the Capital Normal University. His research interests focus on Query Understanding, Sentiment Analysis, Personalized Search, Social Computing.

He has published his important work on the Proceedings of the ACM SIGIR Conference, International Conference on Computational Linguistics (COLING), Journal of Chinese Information Processing and so on.

Dr. Jingli Lu received a doctoral degree in Computer Science from Massey University, New Zealand. She is currently a Software Engineer at Agresearch Ltd, New Zealand. Her research interests include Data Mining and Machine Learning, Fuzzy Set Theory, and Speech Signal Processing. She has published several papers in highly-rankly international journals and conferences.