Combining Domain-Specific Sentiment Lexicon with Hownet for Chinese Sentiment Analysis

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

***Abstract*—Various sentiment analysis approaches have been proposed but little has been done for Chinese documents. In this paper, a small set of Chinese adjective lexicon has been constructed in which each word is context-independent and is labeled with polarity intensity manually. Meanwhile, we introduce a framework which combines context-sensitive sentiment lexicon and the existing sentiment word list such as "Hownet" to enhance the domain specific opinion data analyzing performance. An evaluation experiment which uses the real online product reviews as input also discussed in this paper.**

***Index Terms*—context-sensitive sentiment lexicon, sentiment analysis, polarity classification**

I. INTRODUCTION

Following the emergence of web 2.0 technologies, opinion data which embedded in blogs and online reviews have grown explosively in recent years. This has also brought new opportunities and challenges to opinion mining and sentiment analysis technology. Various approaches have been proposed by different researchers, such as opinion retrieval [1, 14], sentiment classification, sentiment summarization [2], etc. Among those approaches, sentiment lexicon plays a significant role.

Many researches are dedicated to construct a general-purpose sentiment lexicon which can be used to any domain [3-4]. Such resources provide information about the semantic orientation of single words or whole phrases. E.g. an entry is associated with categories like positive, negative or neutral appraisal. However, a general-purpose sentiment lexicon cannot get the best result of the semantic orientation of a document, since the sentiment of some words is context-aware and domain-sensitive. For example, "unpredictable" is likely to be positive in movie reviews, while being negative in laptop reviews. At the same time, a number of words represent affirmative sentiment and are domain-independent. For instance, not matter in what domain, "good" and "bad" always express the polarity of "positive" and "negative" respectively. Currently, most researches of sentiment analysis have been focused on English documents, and little study has been conducted on Chinese sentiment analysis.

In this paper, we will construct a domain specific sentiment lexicon which takes Chinese online product reviews as our input. According to the characteristics of a review document, it always contains two kinds of words:context-dependent and context-independent, which we have mentioned before. The task of construction a sentiment lexicon has been split into to two steps. Firstly, we construct a small word list of sentiment-conveying terms and manually label each word with a polarity score. Secondly, we build a sentiment lexicon exploiting the word list we have built in the first step and the corpus statistics method. Each entry in the lexicon is a combination of product aspect, opinion word and polarity score (i.e. keyboard#good:1). In particular, the corpus statistics method is based on the intrinsic orientation of advantage and disadvantage texts available in many product reviews. We propose an approach that utilizes our own sentiment lexicon and the existing thesaurus (i.e. "Hownet") to get the sentiment orientation of reviews.

The rest of the paper is organized as following: in the next section, previous work on sentiment lexicon induction and semantic classification is described in a more detail way. Our own method is detailed in section 3. Section 4 shows the experiment results. We conclude this paper in section 5.

II. RELATED WORK

Sentiment analysis is an attempt to deal with evaluative aspects of text, which has drawn increasingly attention recently. In sentiment analysis, one of the fundamental tasks is to identify whether the given text expresses positive or negative orientation; and sentiment lexicon plays an essential role to this task. In order to build such lexicon, many researchers have investigated various kinds of methods. So far, literatures on sentiment polarity lexicon induction can be broadly classified into two categories: One is based on the thesaurus and the second type is based on corpus.

Thesaurus based approach utilizes synonyms or glosses to determine the polarity of words. Kamps et al. [5] propose to measure the relative distance between unclassified words and seed words. They build lexical network by using the synonym relations available in WordNet. This method relies on a hypothesis that synonyms have the same polarity. Hu and Liu [6] extend Kamps' method. They use not only synonyms but also antonyms to measure the distance between seed words and unclassified words. Esuli and Sebastiani [7] determined the orientation of subjective terms utilizing glosses, i.e. the definitions of these terms given in online dictionaries. Another relevant example is the recent work

doi:10.4304/jcp.8.4.878-883

made by Mihalcea et al. [8] on multilingual sentiment analysis which using cross-lingual projections. This is achieved by using bridge resources like dictionaries and parallel corpora to build sentence subjectivity classifiers for the target language (Romanian).

Corpus based approach deduce polarity by assuming that sentiment words co-occur with each others are likely to convey the same polarity. There are numerous studies in this field. Hatzivassiloglou and McKeown [9] infer polarity of words by exploiting constraints on conjoined adjectives. For example, two words linked by "but" are most likely to be in opposite polarities, while conjunctions like "and" are proofs for words in the same polarity. Turney and Littman [10] determine a semantic orientation of a phrase by comparing whether it has a greater tendency to co-occur with positive seed words or with negative seed words as measured by point-wise mutual information. The mutual information is estimated by the number of hits returned by a search engine. While the method is general applicable to several domains and languages, it ignores the context-aware characteristic of a word. Kaji and Kitsuregawa [11] propose a method of building sentiment lexicons for Japanese by using HTML layout structure. Besides being very refined to Japanese, excessive dependence on HTML structure makes their method brittle. A.,et al. [17] analyze the sentiment orientation of news utilizing lexicon.

More importantly, many works have considere the context-dependent problem. Few of them investigate the issue that even the same word in the same field may express different orientation. A few studies e.g. JurgenBroß and HeikoEhrig [13, 16] try to generate the semantic orientation of words dependent on aspects. While most of them rely on a single source of information, which is inadequate demonstrated in this paper.

III. Proposed Framework of Context-aware Sentiment Lexicon and Hownet

Our goal is to induce a context-aware sentiment lexicon to identify the semantic orientation expressed in a text. Now we give several definitions used in our method.

Definitions

- Product aspects (or features): we construct an aspect set $A = \{a_1, a_2, a_3, ...a_n\}$; a_i represents a component of a product (such as keyboard or screen) or a product itself (e.g. laptop, computer) or some features depicting a product (e.g. price, speed).
- Adjective set: an adjective word list W_a refers to a set of sentiment-conveying terms that are context-independent. Each word in the list is followed by a sentiment score. And the absolute value of the score indicates the sentiment intensity of each word.
- Context-dependent sentiment lexicon(CSL): it contains a dictionary of sentiment terms conditioned on different product aspects of the given domain. Each entry in the lexicon is a pair of product aspect a_i and opinion word w. Each pair is assigned to a sentiment score representing the polarity it is expressing.

Method details

We take the product reviews (see Figure 1) which is divided into advantage texts, disadvantage texts and user feeling texts by the users as our model's input. One of our major outputs is the aspect based sentiment lexicon. Our method mainly consists of four major steps: data collecting, data preprocessing, building the context-sensitive sentiment lexicon and opinion analysis. We describe each of them in a detail way.

优点 (Advantage): 电脑不错,性价比高！(The computer is good and cost-effective is high. .)
缺点 (Disadvantage): 喇叭声音小了点(Trumpet sound is a little bit small.)
使用心得 (User feeling): 便宜,方便,散热不错 (Cheap, convenient ,thermal dissipation is good)

Fig.1. A review of laptop on 360buy.com

A. Data Collecting

The goal of the data collecting phase is to construct a raw dataset of product reviews for building the subjectivity-lexicon. In this phase, reviews are crawled from 360buy.com. Thus, irrelevant posts (ads or duplicate reviews) are filtered out by a rule-based scheme; those remaining posts are considered as relevant reviews, denoted by $R = \{r_1, r_2, r_3, ...r_n\}$. Each r_i is segmented into several clauses using punctuations such as "."."!" ","," as signals. We denote these clauses by $C = \{c_1, c_2, c_3, ...c_k\}$.

B. Data Preprocessing

In this stage, firstly, we preprocess each review with a part-of-speech tagger (ICTCLAS). By analyzing the tag results, we realize that the part-of-speech of product aspects is always substantive and the opinion words related to aspects are usually expressed by adjective. Another explicit feature of a review is that some commonly used opinion words , a large part of which are context-independent e. g. "good", "bad", are more frequently used than some domain-specific words. Secondly, we induce the candidate aspect set and adjective set by filtering out the frequently appeared nouns and adjectives. Lastly, with domain expert's intervention, the final aspect set A is achieved. Similarity, to the adjective set W_a, we manually filter out noise and label each word with a score in the interval [-1, 1].

C. Building the Context-sensitive Sentiment Lexicon

In this step, at first, we extract candidate opinion words to be paired with aspects. Specifically, we recognize the

product aspect involved in the aspect set and the mapping opinion word involved in the same clause, which is based on the POS attributes (noun and adjective). And we regard the mapping result as a candidate pair. If one clause has been identified with more than one aspect or adjectives, the potential opinion word will be mapped with its nearest aspect. An example sentence is as follows:

笔记本/n 音质/n 稍/d 差/a (The sound quality /n of the laptop/n is slightly worse/a.), where /n ,/d and /a represent noun, adverb and adjective respectively. And there are two nouns in this sentence. In line with the mapping rules, the pair (音质(the sound quality)#差(worse)) is extracted from the example sentence.

Then a value representing polarity intensity will be assigned to each pair. We deal with it in two ways. One is that, in case, the term expressing sentiment in each pair is involved in W_a we have constructed in the previous step, we assign the corresponding value of the word directly to the pair. The other case, if W_a does not contain the term, we compute the sentiment score of the pair utilizing statistical method. This method is based on an assumption that opinion words appearing in advantage text (disadvantage texts) tend to have positive orientation (negative orientation). For each entry, we count its frequency in advantage texts and disadvantage texts.

Negation words such as “no”, “not”, “never” reverse the sentiment of the opinion word in a clause. Therefore, we process them separately. We manually construct a negative word list including commonly used negative words (e.g. “no”, “not”). It is recognized most negative words are adverbs and an adverb always co-occurs with an adjective. Making use of this characteristic, we map the negative word with the opinion word nearest to it. If there are two or more than two negative words in a clause, it is difficult to determine the sentiment contained in the clause. In this case, we ignore the negative indicators and regard this clause as affirmative.

键盘/n 不/d 是/v 很/d 好/a (The keyboard /n is/v not/d very/d well/a.)

There are a noun and an adjective in the above example. Based on our method, the pair 键盘#好 (keyboard#good) is extracted from the sentence. At the same time, the sentence also includes a negative word 不 (not), which we need to deal with using the method introduced above. Then we get a new pair 键盘#不好 (keyboard#not good). If the pair 键盘#好(keyboard#good) is a member of CSL, we can achieve the sentiment score of 键盘#不好 (keyboard#not good) by reversing the value of its positive expression. Otherwise, the proposed statistical method is used to calculate the polarity of the pair with negation term. Specifically, we can count the frequency of the pair (keyboard#not good) in advantage texts (disadvantage texts).

After that, we propose a method to compute the sentiment score expressing by an entry. The function is defined as follows:

$$S(e) = \frac{1}{\chi}(\log(p_{pos} + 1) - \log(p_{neg} + 1)) \quad (1)$$

$$P_{pos} = \frac{F_A}{F} \quad (2)$$

Where $S(e)$ indicates the sentiment intensity of an entry; the higher the absolute value of the sentiment estimation score, the more intense the sentiment expressing by the entry. The parameter χ which is used to normalize $S(e)$ is a constant. The value $F = F_A + F_D$ represents the frequency of an entry appeared in advantage and disadvantage text, where F_A (F_D) is the number of an entry in advantage texts (disadvantage texts). The term P_{pos} (P_{neg}) is the estimated probability that an entry is rated positive (negative).

D. Opinion Analysis

To predict the polarity of an unseen opinionated document, the constructed domain-specific sentiment lexicon is referred to. We estimate the polarity of an opinionated document utilize the function above:

$$\mathrm{SP}(\mathrm{D}) = \lambda \sum_{\mathrm{e} \in W_a} v_e + (1-\lambda) \sum_{\mathrm{o} \in S_e} v_o \quad (3)$$

Where $\mathrm{SP}(\mathrm{D})$, which is generated by the system constructed sentiment lexicon W_a and the sentiment word list involved in Hownet, represents the sentiment polarity of a document D. We use parameter λ to indicate the weight of W_a in sentiment classification task. e is an entry contained in W_a and its corresponding sentiment value is v_e .The parameter o represents sentiment conveying-words involved in document D and the Hownet sentiment word list(Se); we define v_o as follows:

$$V_o = \begin{cases} 1 & if\ \mathrm{o} \in positive\ wordlist\ in\ Hownet \\ -1 & if\ o \in negative\ wordlist\ in\ Hownet \\ 0 & others \end{cases} \quad (4)$$

If $\mathrm{SP}(\mathrm{D}) > \phi$, the pair is considered to be positive. Analogously, hypotheses are formulated to decide whether a text exhibits negative orientation. Another case that neither positive nor negative orientation can be assigned to the text, it is classified as being neutral. The parameters χ , λ and ϕ are empirically established based on the training corpus of a specific domain. For the experiment in this paper, we adopted the following values: $\chi = 0.69$, $\lambda = 0.70$, $\phi = 0.02$.

IV. Experiments And Results Analysis

A. Evaluation of Lexicon Quality

There is no existing data set available to evaluate the quality of a constructed context-dependent sentiment lexicon, which is in the form of a sentiment score

assigned to each aspect-opinion pair. So, we apply our method to the below-mentioned corpus and manually annotate each extracted pair with its true class (positive, negative, neutral).Similar to previous studies [12, 15], we retrieve real-world opinionated documents from the Web to build our evaluation data set. More specifically, 3765 laptop reviews are downloaded from 360buy.com, invoking our crawler programs. Applying the proposed method on the corpus, we are able to construct the aspect set (32 aspects) and the context-independent sentiment word list (60 opinion words). Then, we can extract a total of 2163 unique pairs, of which 1256 are sentiment relevant. Thus, 841 pairs are considered having positive orientation, 415 being negatively connoted utilizing the proposed method. In Table I, we present part of results of the context-independent word list. The results of our methods on entry labeling are shown in Table II. The time complexity of building such a lexicon is n^3 (n represents the number of clauses in the training dataset).

TABLE II.
PART OF SYSTEM DISCOVERED DOMAIN-INDEPENDENT SENTIMENTS

正向极性(Positive polarity)	负向极性(negative polarity)
好（good）:1	不好(not good):-1
不错(not bad):1	差(bad) :-1
便宜(cheap):1	烂(poor) :-1
划算(cost-effective):1	不行(not work very well):-1
正常（normal）:0.5	麻烦(inconvenient):-1
实用(practical):1	贵(expensive):-1
完美(prefect):1	一般(so so):-0.5
理想(ideal):1	笨重(heavy):-1
强大(powerful):1	坏(awful):-1
漂亮(beautiful):1	失望(disappoint):-1
大方(natural):1	慢(slow):-1
实惠(affordable):1	粗糙(coarse):-1
好看(good-looking):1	脆弱(fragile):-1
清晰(clear):1	丑(ugly):-1
给力(awesome):1	不足(inadequate):-1
可靠(reliable):1	普通(ordinary):-1

Compared with manually annotated result, the precision of our method is 90.75%. The good results of this experiment are mainly due to the follow reasons. One is that, we exploit the inherent features of a review (advantage text, disadvantage text) and syntactic rules (e.g. relationship between noun and adjective). The statistical method plays a significant role in filtering out irrelevant pairs, which is another reason.

B. Evaluation of Sentiment Classification using Lexicon and Hownet Sentiment Word List

In this experiment, our purpose is to examine the performance of the proposed framework (SP), which combines the information coming from the constructed lexicon and the sentiment word list in Hownet, in sentiment classification task. The output of the framework is a sentiment score which is used to determine the sentiment orientation of the given text. To execute this test, 1892 laptop reviews crawled from web are used as the evaluation set. Each review is manually labeled with a sentiment polarity (positive, negative, neutral) by the annotators who have annotated the orientation of the pair contained in our lexicon. This enables us to evaluate the sentiment classification performance of different methods. As comparison, we consider the following baselines for classifying the sentiment orientation of documents: utilizing Hownet sentiment word list only; using context-aware sentiment lexicon (CSL) only.

TABLE I.
SOME ENTRIES OF SENTIMENT LEXICON FOUND IN COMPUTER CORPORA

Aspects	Entries
显卡 (graphics card)	显卡#给力(graphics card#awesome):1.0
	显卡#不好((graphics card#not good):-1.0
性价比 (performance ratio)	性价比#一般(performance ratio#so so):-0.5
	性价比#好(performance ratio# good):1.0
内存 (memory)	内存#小(memory#small):-0.652
	内存#大(memory#large):0.923
处理器 (processor)	处理器#不错(processor#not bad):1.0
	处理器#一般(processor#so so) :-0.5
键盘 (keyboard)	键盘#结实(keyboard#solid):1.0
	键盘#软(keyboard#weak):-0.964
电脑 (computer)	电脑#好看(computer#good-looking):1.0
	电脑#厚(computer#thick) :-0.323
外观 (appearance)	外观#大方(appearance#natural):1.0
	外观#厚重(appearance#heavy) :-0.5
手感 (feel)	手感#好(feel#good) :1.0
	手感#差(feel#bad) :-1.0
速度 (speed)	速度#快(speed#fast) :1.0
	速度#慢(speed#slow) :-1.0
摄像头 (camera)	摄像头#清晰(camera#clear) :1.0
	摄像头#差劲(camera#poor):-0.588
屏幕 (screen)	屏幕#亮(screen#bright) :1.0
	屏幕#黑(screen#black) :-0.488
硬盘 (hard disk)	硬盘#完美(hard disk#perfect):1.0
	硬盘#小(hard disk#small):-0.485
音响 (sound)	音响#强大(soun#powerful):1.0
	音响#差(sound#bad):-1.0
做工 (workmanship)	做工#精致(workmanship#exquisite) :1.0
	做工#粗糙(workmanship#coarse) :1.0

The outputs of different methods can be evaluated by:

$$Precision = avg(Prec(pos) + Prec(neg) + Prec(neu)) \quad (5)$$

$$\text{Recall} = avg(\text{Recall}(pos) + \text{Recall}(neg) + \text{Recall}(neu)) \quad (6)$$

$$F - measure = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

$$\text{Prec}(pos) = \frac{N_{Pcor}}{N_{Psys}} \quad (8)$$

$$\text{Recall}(pos) = \frac{N_{Pcor}}{N_{Pall}} \quad (9)$$

Where Precision is the average precision of our system generated results in different classes (positive class, negative class, neutral class). N_{Pcor} is the amount of correct positive reviews compared with manually labeled results. N_{Psys} is the number of positive reviews detected by our method. Similarity, Recall represents the average recall. N_{Pall} is the number of the manually annotated positive documents in the evaluation texts. The results of the experiment are reported in Table III.

We summarize the results in Table III and highlight in bold font the best performance under each measure. As a whole, the proposed method is quite efficient to process opinionated documents. The improvement is mainly due to the increasing in the input information. Exploiting thesaurus only, a lot of sentiment message based on aspect is ignored, e.g. “The screen is small”, has a negative orientation, whereas utilizing thesaurus cannot identify it because the word “small” does not have definite orientation. Similarity, only considering the information contained in the constructed lexicon is far more from enough. For example,” very good!” has positive orientation apparently, but there is no mapping aspect to the opinion word “good”. Therefore, our method which takes advantage of both gets better results.

TABLE III.
SENTIMENT CLASSIFICATION PERFORMANCE OF DIFFERENT METHODS

Method	Precision	Recall	F-Measure
Hownet only	0.5024	0.4628	0.4729
CSL only	0.6853	0.6296	0.6563
SP	**0.7502**	**0.7018**	**0.7252**

V. CONCLUSIONS AND FUTURE WORK

In this paper, we explore to construct a domain-specific sentiment lexicon to be applied in an aspect-based review mining scenario. And we formulate a framework to combine the constructed lexicon and the existing sentiment word list. We have demonstrated that our method can learn new context-aware words and aspect-dependent sentiment. Thus, our method could achieve a better result in sentiment classification tasks. Experimental results demonstrate the feasibility of our approach.

As future work, we can exploit other kinds of useful signals such as synonym and antonym relationship between terms. How to identify the accurate polarity of the entries in our lexicon is also taking into consideration. We intend to utilize SentiWordNet and syntactic parsing to solve this problem. We also plan to evaluate the effectiveness of our context-aware sentiment lexicon in other sentiment related applications, such as opinion retrieval and opinion summarization. Another interesting future work is to study how to adapt our domain-specific lexicon to other domains.

ACKNOWLEDGMENT

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

REFERENCES

[1] S.-H. Na, Y. Lee, S.-H. Nam, and J.-H. Lee, “Improving opinion retrieval based on query-specific c sentiment lexicon,” In ECIR '09, pp. 734-738, Berlin, Heidelberg, 2009.

[2] I. Titov and R. McDonald, “A joint model of text and aspect ratings for sentiment summarization,” in Proceedings of ACL-08: HLT. Columbus,OH: Association for Computational Linguistics, pp. 308–316 June 2008.

[3] Taboada, M., Brooke, J, Tofiloski, M., Voll, K.D., Stede, M, “Lexicon-based methods for sentiment analysis,” In Proceedings of Computational Linguistics, pp. 267–307, 2011.

[4] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, “Sentiful: Generating a reliable lexicon for sentiment analysis,” In ACII, pp. 1 -6, 2009.

[5] J. Kamps, M. Marx, R. J. Mokken, and M. de Rijke, "Using WordNet to measure semantic orientation of adjectives," in Proceedings of LREC, 2004.

[6] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in KDD’04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: ACM, pp. 168–177, 2004.

[7] Esuli, A. &Sebastiani, F, “Determining the semantic orientation of terms through gloss classification,” In Proceedings of the 14th ACM international conference on Information and knowledge management, Bremen, Germany, 2005.

[8] RadaMihalcea, Carmen Banea, and JanyceWiebe. “Learning multilingual subjective language via cross-lingual projections,” In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pp. 976–983, 2007.

[9] Vasileios Hatzivassiloglou and KathleenMcKeown. “Predicting the semantic orientation of adjectives,” In Proceedings of the ACL, pp. 174–181, 1997.

[10] P. D. Turney and M. L. Littman, “Measuring praise and criticism: Inference of semantic orientation from association,” In Proceedings ACM Transactions on Information Systems (TOIS), 2003, pp. 315–346.

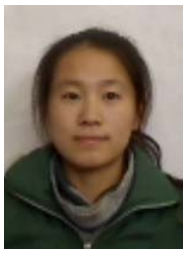
[11] N. Kaji and M. Kitsuregawa, “Building lexicon for sentiment analysis from massive collection of HTML documents,” In Proceedings of EMNLP-CoNLL 2007. Prague, Czech Republic: Association for Computational Linguistics, pp. 1075–1083, 2007.

[12] H. Wang, Y. Lu, and C. Zhai, “Latent aspect rating analysis on review text data: a rating regression approach,” In Proceedings of KDD '10, New York, NY, USA, pp.783-792, 2010.
[13] JurgenBroß and HeikoEhrig, “Generate a context-aware sentiment lexicon for sentiment analysis,” In Proceedings of IEEE '10, pp. 435-439, 2010.
[14] V. Jijkoun, M. de Rijke, and W. Weerkamp,“Generating focused topic-specific sentiment lexicons,”In Proceedings of ACL '10, pp. 585-594, 2010.
[15] Y. Choi and C. Cardie., “Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classication,” In Proceedings of EMNLP '09, pp. 590-598, 2009.
[16] Yue Lu, Malu Castellanos, Umeshwar Dayal, ChengXiang Zhai, “Automatic construction of a context-aware sentiment lexicon: an optimization approach,” Proceedings of the 20th international conference on World Wide Web, 2011.
[17] A. Moreo, M. Romero, J.L. Castro, J.M. Zurita, “Lexicon-based Comments-oriented News Sentiment Analyzer system”,Expert Systems with Applications, 2012.

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.



Mengyun Lei is studying for a master's degree at CapitalNormalUniversity. Now, her research interest includes Opinion Mining, Text Classification, Intelligent Information Processing Systems, and Data Mining.

Dr. Hanshi Wang holds a PhD in Computer Application from the Beijing Institute of Technology, China. He is currently a lecturer at the CapitalNormalUniversity. 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.