Sentiment Analysis in Chinese Web Discussion Forums

Ming Liu*, Weiwei Xu, Huiran Liu, Jiaqin Zheng
Southwest University, Beibei District, Chongqing, China, 400715.

* Corresponding author. Tel.: +86 15178719397; email: mingliu@swu.edu.cn
Manuscript submitted February 10, 2015; accepted March 22, 2015.
doi: 10.17706/ijeeee.2015.5.1.47-54

Abstract: Online discussions serve as a bridge between face-to-face and virtual lessons in the distance education. Previous researches have investigated the impact of student participation in online discussions on learning performance. However, student opinions expressed in the posts are also good indicators to student learning. The recent achievements of the natural language processing technologies, especially sentiment analysis in Chinese makes it possible to automatically detect the polarity of each course post. In this study, we collected online posts from 62 students in a distance education course program and used machine learning algorithms to build a sentiment classifier. The features were extracted from a recent developed Chinese affective lexical resource. Finally, we examined the impact of the post opinion on the student final grade. As a result, it is found that the decision tree classifier got the highest Marco-average F_1-score (0.693) and the correlation between the student post sentiments and final grade is moderate.

Key words: Sentiment analysis, student forum posts, language technologies.

1. Introduction

According to Kearsley [1], the most significant application of computer-mediated communication in E-learning environments is discussion forum. Online discussion forums provide a way for students to extend the classroom discussions. They provide better cognitive and exploratory learning [2], increase student-to-student discussion and cooperation [3], and upgrade critical thinking skills [4]. In addition, Palmer *et al.* [5] investigated the impact of participation in online discussion on student performance. Their study results indicated that assessing online discussion positively impacted students' participation and final grade. Ravichandran and Purushothman [6] extended this work to building a regression model which could use forum score to predict the final grade.

The opinions embedded in students' posts are useful for dropout rate prediction and course evaluation. Wen *et al.* [7] explored mining collective sentiment from forum posts in a Massive Open Online Course (MOOC) in order to monitor students' trending opinions towards the course and major course tools, such as lecture and peer-assessment. They observed a correlation between sentiment ratio measured based on daily forum posts and the number of students who dropped out each day. In addition, analyzing the forum posts, we can infer important information about the attitudes prior to the post-course surveys [7].

Up till now, many researches have been conducted on English document sentiment classification. These researches have fallen into two categories: machine learning based approach [8], [9] and semantic orientation approach [10]-[12]. Machine learning based approach focuses on training a sentiment classifier based on occurrence frequencies of various words in the documents. Pang *et al*. [9] applied three classification algorithms—Naïve Bayes, Maximum Entropy and Support Vector Machines—to classify movie reviews into positive and negative. On the other hand, the semantic orientation approach first classifies words into two categories, positive and negative, and then counts an overall positive/negative score for the text. If a document contains more positive than negative words, it is regarded as positive; otherwise, negative. In general, the study results showed that machine learning based approach outperformed the semantic orientation approach [13]. More recently, Socher *et al*. [14] proposed recursive neural tensor network model trained on the new sentiment Treebank, which outperformed previous methods.

For the existing Chinese sentiment analysis research, however, little effort has been put into the forum domain [15], while most papers focus on investigating opinions from online review sites [16], [17] and Weibo (Chinese Twitter) [18], [19]. Zhao *et al*. [18] used machine learning approach to train a statistical classifier to classify Chinese tweets into four categories of emotions (angry, disgusting, joyful and sad). The study results showed the best precision is 64.3% with Naïve Bayes. Yuan *et al*. [19] also employed machine learning based approach to classify Chinese tweets to six emotions defined by Ekman [20]. The best accuracy reached 80% with SVM model.

In this paper, we present an empirical study of sentiment classification on Chinese online discussion forums in a distance education system. The contribution of this paper is described as follows:

1) We use recent developed Chinese affective lexical resource to extract features;
2) We evaluate three machine learning algorithms in a real dataset collected from a distance education course program in China;
3) We examine the impact of student opinions posted in the discussion forum to the final grade.

The rest of this paper is constructed as follows: Section 2 presents related work on affective lexical resources. Text preprocessing and feature selection are described in Section 3. Experimental results are given in Section 4. Finally, Section 5 concludes this paper.

2. Related Work

Emotion lexicons have a great impact on the sentiment analysis. In this section, we briefly review related work on Chinese affective lexical resource.

Affective lexical resource plays an important role in feature extraction in sentiment analysis. In English, WordNet-Affect [21] is an extension of WordNet-domains that label synsets with representing affective concepts. Synset is the synonym set that represents a sense or a concept in the WordNet lexicon. 2,874 synsets and 4,787 words are annotated in the WordNet-Affects. Lexical information including part-of-speech, synonyms, antonyms and definitions are manually added. The SentiWordNet [22] is another lexical resource developed for supporting sentiment classification. It scores each synset in the WordNet along three sentiment dimensions: positivity, negativity and neutrality. Those pre-developed emotional dictionaries can be applied to emotion classification directly.

In Chinese, the most used lexical resource is HowNet [23]. HowNet is a Chinese network of lexical resource that contains conceptual and attribute relations among words. It only contains 6 sub-files for sentiment analysis, such as positive and negative feeling, positive and negative sentiment, opinion and degree. In addition, it has hierarchical relations among lexicons. More recently, Cheng and Lin [24] manually developed an affective lexical resource which contains 27,466 affective Chinese lexicons including words and phrases. Each lexical information includes part-of-speech, emotion category, polarity and affective degree. The emotion taxonomy is an extension of 6 basic emotions proposed by Ekman [20] and defines 7 main emotion categories (joy, good, anger, sadness, fear, disgust, surprise) and 21 emotion sub-categories.

Table 1 shows the emotion categories used by Cheng *et al*. In our study, we use machine learning based approach to build the post polarity classifier. The features are extracted based on the Chinese affective lexical resource proposed by Cheng *et al.* [24].

Table 1: Chinese Emotion Categories

Main Category	Sub-Categories
高兴 (joy)	快乐(happiness) 安心(relief)
好(good)	尊敬(respect) 赞扬(praise) 相信(trust) 喜爱(fondness) 祝愿(wish)
愤怒(anger)	愤怒(anger)
悲哀(sadness)	悲伤(sadness) 失望(disappointment) 疚(regret) 思(miss)
害怕(fear)	慌(panic) 恐惧(fear) 害羞(shame)
厌恶(disgust)	烦闷(irritation) 憎恶(hatred) 贬责(criticism) 妒忌(envy) 怀疑(suspicion)
惊奇(surprise)	惊奇(surprise)

3. Text Pre-processing

In order to make the textual data available for machine learning algorithm, we need to transform the textual dataset to a feature table. Unlike English, Chinese does not have space between words. Therefore, the Chinese text pre-processing includes two main steps: word segmentation & parsing and feature extraction.

Step 1: Word Segmentation & Parsing. There are some off-the-shelf Chinese word splitters publically available, such as FudanNLP (https://code.google.com/p/fudannlp/), Institute of Computing Technology, Chinese Lexical Analysis System (ICTCLAS) designed by the Chinese Academy of Social Sciences (http://ictclas.nlpir.org/). These splitters not only segment word, but also perform parsing. In the study, we used the FudanNLP tool to process each post and got each word in the post with a part of speech.

Step 2: Feature Extraction. Instead of traditional Bag of Words model for sentiment analysis, we developed 13 fine-grained feature set. Based on Chinese affective lexical resource [24], nine numeric type features were defined: NumofJoy, NumofGood, NumofAnger, NumofSadness, NumofFear, NumofDisgust, NumofSurprise, NumofPositive and NumofNegative. These features would capture the emotional words appearing in each post. From Step 1, we first obtained the word with a part of speech. Then we looked it up in the lexical resource to see if it matched one of the emotion categories.

In addition, we extracted features from emotion icons (握手|handshake,拥抱|hug), punctuation symbols(! Or ?), word of gratitude (e.g. 谢谢|thanks) and word of encourage (加油|fighting) features.

4. Study

4.1. Participants and Procedure

A total number of 62 university students participated in this study. The participants' age ranged from 21 to 40 (M: 27, SD: 4.54) and there were 23 males and 39 females. The student participants came from different disciplines, including engineering (21), education (12), commerce (14), English (5) and others. The participants, all volunteers, signed an informal consent form approved by human research ethics committee.

The participants were enrolled in a distance education program at a university in China. The distance education institute offers 49 majors to around 82,000 students all over the country. Like Cousera, the distance education system is the only learning channel and provides various types of learning materials, such as video and lecture notes. More importantly, the online discussion forum is frequently used by students. In this study, we randomly chose 62 students and analyzed their posts on the forum.

Those 62 students posted 6,684 posts in one semester. Two human annotators were asked to independently annotate the polarity of each post. As a result, the Cohen's Kappa coefficient is 0.57, indicating moderate reliability. During the annotation process, the main problem was the ambiguous posts. For example, 我真是太聪明了(I am so smart). It could mean "I am really smart" or "I am just being sarcastic". Without the context it is difficult to determine the polarity of this post.

Table 2 shows the distribution of each category. As expected, most posts are neutral. The number of positive posts is greater than that of negative posts.

Table 2: Student Posts Polarity Distribution

Category	Number of posts	Example
Positive	1902	互相帮助 一起努力 (Help each other and work together.)
Negative	1265	心理学考什么 一片茫然 感觉什么也不知道了 (I have no idea about how to prepare for the final exam in psychology?)
Neutral	3519	计算机统考难不难啊！具体要考些什么啊(Is it difficult to pass the test in the computer course? What the exam questions would be?)

4.2. Sentiment Classification Result

In this study, we used Weka [25] , a machine learning software toolkit, as our evaluation platform. We tested the post polarity classification performance among three classification algorithms: Naïve Bayes (NB), J48 (Decision Tree) and Sequential Minimal Optimization (SMO). The Weka implementations of SVM, with a complexity of 2.0 and Polynomial kernel function with an exponent parameter 2 (The parameters were obtained from our experiment in order to get the best result on SVM).

We used balanced F_1-score, precision and recall to measure the classifier's performance. The F_1 of each class is computed by the following formula:

$$f_1 = \frac{2 \times precision \times recall}{precision + recall} \quad (1)$$

The precision for a class is the number of true positives (i.e. the number of positive posts correctly labeled as belonging to the positive class) divided by the total number of elements labeled as belonging to the positive class, while the recall for a class is the number of true positives divided by the total number of elements that actually belong to the positive class.

Table 3 shows the sentiment classifier's performance and how well each class was predicated by each classifier with 10-fold cross validation results. It was found that J48 outperformed SMO and Naïve Bayes in the macro-average F_1-score (0.693). Although Naïve Bayes got a good precision (0.934) and SMO a good recall (0.955) in Neutral category, they had a poor performance in classifying Negative posts. J48 got a more reliable performance in each category.

Table 3: The Average Scores of Three Classifiers

Classifier	Class	Precision	Recall	F_1-score
Naïve Bayes	Negative	0.289	0.846	0.431
	Neutral	0.934	0.159	0.272
	Positive	0.583	0.795	0.672
	Avg.	0.718	0.467	0.419
SMO	Negative	0.535	0.172	0.261
	Neutral	0.615	0.955	0.748
	Positive	0.886	0.364	0.516
	Avg.	0.682	0.644	0.595
J48	Negative	0.631	0.307	0.413
	Neutral	0.684	0.894	0.775
	Positive	0.824	0.619	0.707
	Avg.	0.717	0.711	0.693

We investigated some issues which could influence the classifiers' performance.

Textbook posts: some posts containing sentiment words came from the textbooks or other learning materials. This would lead to poor precision. For example, 孙中山虑误期泄漏消息 (Sun Yat-Sen worries that the release of the information is delayed).

Informal expression: it includes dialect, verbal expression and internet vocabulary. This informal expression would lead to poor recall. For example, in this sentence, 这是看书要把人看疯了的节奏 妈蛋 (Reading the book makes me crazy). "看疯" and "妈蛋" containing negative polarity are informal expression, which are not listed in the affective lexical resource.

4.3. Correlation between Post Polarity and Final Grade

In order to evaluate the impact of the posts on student final scores, we calculated the average post polarity of each student enrolled in a course by the following formula：

$$\frac{1}{n}\sum_{i}^{n} a_i \tag{2}$$

where n is the number of post a student posted in a course and a_i is the polarity of each post. In this study, the label for Positive post is 1 where the label for negative post is -1. Neutral post is 0.

It is observed that Pearson coefficient is 0.547 (n=126) between the average post polarity of each student and the student's final score, illustrating a moderate correlation. This result indicated that the opinions expressed in the forum would moderately correlate the final score. One of the problems for causing the low coefficient is the repeated posts. The same texts containing sentiment words have been posted multiple times. For example, 为了这20分真的不容易啊！终于看到希望就在前方 (It is not easy to get these 20 marks! I finally see hope ahead.). This message containing positive words were posted 8 times.

5. Conclusion

With the increase of the use of distance education system, online discussion forums have gained popularity in capturing the communication process between teacher-student, student-student, and student-teacher. Earlier researches showed the participation in online forums could lead to broader and deeper participation in group activities and examined the impact of discussion forums with respect to student final scores. In this paper, we investigated the sentiment expressed in course forum posts in a distance education system in China and found that the sentiment expressed in course forum posts were moderately correlated to the student final scores. In addition, three machine learning algorithms were evaluated in building a sentiment classifier, where the features were extracted from a recent developed Chinese affective lexical resource. The study results showed that J45 achieved better performance. But, it is still a challenging task to develop a more accurate Chinese sentiment classifier due to the various informal expressions and coped textbook text.

This is an early stage from research due to the limited number of students. In the future work, we will look at more specific study groups and sample over few 100 students. In addition, we will investigate opinion mining in forum posts and examine the student opinion towards particular teaching instruments, such as lecture note, video, assignment or quiz, or teachers' feedback. This tool will be useful to support student post-course survey and help teachers and educational managers to improve teaching methods and strategies.

Acknowledgment

We thanks to the online and continuing education college at the southwest university for providing the dataset. This work is partially supported by Chongqing Social Science Planning Fund Program under grant No. 2014BS123, Fundamental Research Funds for the Central Universities under grant No. XDJK2014A002 and No. XDJK2014C141 and No. SWU114005 in China.

References

[1] Kearsley, G. (2000). *Online Education: Learning and Teaching in Cyberspace*. Belmont, CA: Wadsworth.

[2] Haggerty, N., Schneberger, S., & Carr, P. *Exploring Media Influences on Individual Learning: Implications for Organizational Learning*.

[3] Stodel, E. J., Thompson, T. L. & MacDonald, C. J. (2006). Learners' perspectives on what is missing from online learning: Interpretations through the community of inquiry framework. *International Review of Research in Open and Distance Learning, 7*.

[4] Shapley, P. (2000). On-line education to develop complex reasoning skills in organic chemistry. *Journal of Asynchronous Learning Network, 4*.

[5] Palmer, S., Holt, D. & Bray, S. (2008). Does the discussion help? The impact of a formally assessed online discussion on final student results. *British Journal of Educational Technology, 39*, 847-858.

[6] Ravichandran, P., & Kaur, A. (2013). Impact of discussion forums on the final scores of post graduate

students at Open University Malaysia. *ASEAN Journal of Open Distance Learning, 5*.

[7] Wen, M., Yang, D. & Rosé, C. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? *Proceedings of Educational Data Mining* (pp. 1-8).

[8] Mullen, T., & Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. *Proceedings of Conference on Empirical Methods in Natural Language Processing* (pp. 412-418).

[9] Pang, B., Lee, L. & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)* (pp. 79-86).

[10] Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics* (p. 181).

[11] Turney, P. D., & Littman, M. L. (2013). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems, 21*, 315-345.

[12] Whitelaw, C., Garg, N., & Argamon, S. (2005). Using appraisal groups for sentiment analysis. *Proceedings of the 14th ACM international conference on Information and Knowledge Management - CIKM '05* (p. 625).

[13] Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for Chinese documents. *Expert Systems with Applicationsq, 34*, 2622-2629.

[14] Socher, R. *et al.* (2013). Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Seattle, USA.

[15] Glance, N. *et al.* (2005). Deriving marketing intelligence from online discussion. *Proceeding of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining KDD 05* (pp. 419-428).

[16] Tan, S., Wang, Y. & Cheng, X. (2008). Combining learn-based and lexicon-based techniques for sentiment detection without using labeled examples. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR 08* (p. 743).

[17] Ye, Q., Shi, W. & Li, Y. (2006). Sentiment classification for movie reviews in Chinese by improved semantic oriented approach. *Proceedings of the Annual Hawaii International Conference on System Sciences*.

[18] Zhao, J., *et al.* (2012). Moodlens: An emoticon-based sentiment analysis system for chinese tweets. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1528-1531).

[19] Yuan, Z., & Purver, M. (2012). Predicting emotion labels for chinese microblog texts. *in CEUR Workshop Proceedings* (pp. 40-47).

[20] Ekman, P. (1999). *Basic Emotions*.

[21] Strapparava, C., & Valitutti, A. (2004). WordNet-affect: An affective extension of WordNet. *Proceedings of the 4th International Conference on Language Resources and Evaluation* (pp. 1083-1086).

[22] Esuli, A., & Sebastiani, F. (2006). SENTIWORDNET: A publicly available lexical resource for opinion mining. *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC-06)* (pp. 417-422).

[23] Dong, Z. D. Z., & Dong, Q. D. Q. (2003). HowNet — A hybrid language and knowledge resource. *Proceedings of International Conference on Natural Language Processing and Knowledge Engineering.*

[24] Chen, M., & Lin, H. (2009). *The Construction and Application of Chinese Emotion Word Ontology*. Master

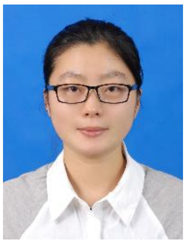
Thesis, Dalian University of Technology.
[25] Hall, M., *et al.* (2009). The WEKA data mining software. *ACM SIGKDD Explorations, 11*, 10-18.

Ming Liu is currently a lecturer at the School of Computer and Information Science, Southwest University, China. He received the PhD in artificial intelligence in education at the School of Electrical and Information Engineering, the University of Sydney, Australia in 2012. He obtained his master degree of information technology from the University of Tasmania, Australia, in 2007.

He worked as a postdoctoral researcher at the University of Sydney in 2013. Dr. Ming Liu has great research interests in learning analytics and intelligent tutoring system. He participated in national and international projects funded by ARC Linkage (Australia), Young and Well CRC, Office of Teaching and Learning, Google and Chinese National Fund. He is an author of over 20 publications papers in prestigious conferences and journals, such as Intelligent Tutoring Systems, IEEE transactions on Learning Technologies, Journal of Educational Technology and Society.

Weiwei Xu holds a doctorate degree in English from Macquarie University, Australia. She acquired her MA degree in English from Newcastle University, UK.

She has been working as a lecturer teaching English academic writing at College of International Studies, Southwest University in China for 8 years.



Huiran Liu comes from the Department of Computer Science and Technology in Southwest University. She is now a senior student majoring in science and technology.

She shows great interests in data mining and mathematical modeling. As for her accomplishments, she has won second prizes in both Chongqing Contest District in China Undergraduate Mathematical Contest in Modeling and American Undergraduate Mathematical Contest in Modeling named Honorable Mention.

Jiaqin Zheng is a senior computer science student at Southwest University. She shows great research interests in sentiment analysis. She got the Meritorious Winner in 2014 Mathematical Contest in Modeling.