

Research on the Intensity of Subjective and Objective Vocabulary in Interactive Text Based on E-Learning

Wansen Wang and Peishen Li

Abstract Based on the text subjective judgment algorithm based on the rough set, we proposed an improved logarithmic linear model and fuzzy set combining the subjective intensity of learning method Chinese words and lexical subjectivity recognition, which is applied in the E-learning interactive text, and achieved better recognition results.

Keywords Log-linear model · Fuzzy set · E-Learning interactive text · Subjectivity intensity

1 Introduction

With the development of network information technology, E-Learning has become an effective form of school education, enterprise training, organization training. However, the traditional E-Learning system lack of emotion generally, in order to increase the emotional functions of E-Learning system, people began to study the emotion of learners studying with E-Learning. Approaches commonly used are: facial expression recognition, text sentiment analysis, speech emotion analysis. In fact, to mine the learners' ideas from academic texts, and then analyzed the

Supported by the National Natural Science Foundation of China under No. 60970052, Beijing National Natural Science Foundation (The Study of Personalized E-learning Community Education based on Emotional Psychology 4112014).

W. Wang (✉) · P. Li
Department of Information Engineering Institute, University of Capital Normal,
Beijing, China
e-mail: wansenw@126.com

P. Li
e-mail: leaps2013@gmail.com

psychological condition is in the premise of academic interactive text subjectivity classification [1].

Subjectivity of Chinese words is a basic problem in text sentiment analysis. Its accuracy will directly affect the follow-ups; it is the basis of the sentiment analysis of the phrase level, sentence level, and paper-level. Although many studies have done [2, 3], the existing analysis methods in dealing with large-scale texts still face the following difficulties: For example, different words in the expression of Opinion may have different subjective intensity, and thus have different effects on subjective analysis of sentences or articles. Moreover, the same words in different language environments may have different subjective intensity, a major problem we are faced is distinguish the subjectivity of words according to the current language environments, but there is less research on the words subjective intensity. Also, a subjective sentence may include two or more subjectivity of the words, but the roles they play to express their opinions are different.

This article firstly introduces the rough set theory for reduction of the text, and secondly extracts corpus from E-learning platform, and the use of rough set theory to reduce; then extracts emotional candidate words and views indicator words, and calculates their subjective weights; Finally, we combine fuzzy set theory, inspecting the impact of the intensity of subjective word on Chinese sentence subjectivity classification.

2 The Text Subjectivity Judgment Based on Rough Set

The so-called subjective text refers to the objective facts described in the text. Its main contents are based on the allegations or arguments, and with one's personal feelings and intentions. A sentence, no matter what form it expressed, as long as the sentence including a subjective component, then it is defined as the subjective sentence. Based on the subjective sentences these features, undoubtedly it is difficult to achieve the purpose of distinct subjective and objective words with methods of analysis of sentence constitutes.

Under the premise of maintaining the same sentence subjectivity, you can arbitrarily change the organization of the sentence, add modifiers. However, regardless of the form of subjective sentences changes the expression of subjective thinking will not change. For example, "I particularly like the movie" change its expression, under the premise of maintaining the same sentence subjectivity, can be said to be: No matter how they evaluate, I like the movie "or" the story of the drama develops more reasonable, but I still like the movie, these three sentences eventually to express personal views are "I like the movie", result of word software process is R R V I R R. So, in the sentence, it can be judged as long as it contains R R V I R R this model, they think this sentence is subjective sentence. This is what we determine the basis of subjective sentence. If a sentence contains the subjective sentences model, then the sentence is subjective sentence; otherwise, this sentence is the objective sentence.

Based on the above ideas, the first thing to do is collecting 1,000 subjective sentences from the E-Learning platform, then extract the structure model of the subjective sentences using word software. To this end, firstly, set these models to a certain threshold, if the threshold is reached, then save this mode; otherwise not retained. In this way, under the premise of ensuring the precision rounding part of the sentence patterns. In the remaining modes may contain redundant elements of sentence or mode of redundancy, for example: model one: R R V I N and model two: R R D V I N, of this case, it is necessary to consider the rough set theory to reduction experimental initial parameters, that is, the use of the knowledge of rough set areas reduction of redundant sentence elements, for example, mode: R R D V I N reduction off “D”, the result is: R R V I N. And then re-use rough set attribute reduction, reduction of redundant attributes; here is the reduction redundant mode, the mode I and mode II reduction for the same mode: R R V I R R. Thus, from the original two modes, and seven sentence elements compare match to the final one mode, and three sentences Comparison of components, and this can improve the efficiency of the implementation of the program to a large extent [4].

Rough Set Theory plays a fundamental role in text for the Reduction and reducing the time cost of the system. The experiments show that either the precision or time cost of the research of the Chinese text based on rough set has a lot of improvements.

3 The Subjective Words Extraction Based on E-Learning

The next three steps are subjective words extracted from the training corpus. First of all, regard the verb as the candidate words of potential views indicator, adjectives and adverbs as underlying emotional candidate words. Then we use log-linear model to calculate the relevance of the words and subjective categories, as the weight of subjective words. Finally, we exclude those who cannot directive as candidate words of advice and emotional according to the weight of words.

In general, a view sentence [5] often contains comments indication words and emotional words to express their views. In Chinese, generally, opinions indication words are some verb, for example, regard, say advocates, these verb and opinions holders jointly published some comments. Emotional information often expressed with some polarity words or phrase, such as the adjective ‘beautiful’, ‘ugly’, the adverb ‘but’, ‘may’. For convenience of presentation, we regard the words related with emotions expressed as emotional words. For log-linear model can be well predict variables and variables, and the degree of correlation between the variables and categories, so we use the log-linear model to predict the weight of words subjectivity, for it can better describe the degree of correlation between words and words, words and subjective categories in our training corpus.

Firstly, let’s calculate the probability and frequency of candidate words in the training corpus. Table 1: The probability and frequency of subjectivity words in

Table 1 Contingency tables of frequencies and probabilities for weighting subjective words

W	C					
	Sub	Obj	\sum_j	Sub	Obj	\sum_j
W_1	n_{11}	n_{12}	$n_{1\cdot}$	p_{11}	p_{12}	$p_{1\cdot}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
W_k	n_{k1}	n_{k2}	$n_{k\cdot}$	p_{k1}	p_{k2}	$p_{k\cdot}$
\sum_i	$n_{\cdot 1}$	$n_{\cdot 2}$	n	$p_{\cdot 1}$	$p_{\cdot 2}$	p
	Frequency table			Probability table		

the training corpus. Here W represents the words in the training corpus, C represents the subjective and objective categories, namely {subjective sentences, the objective sentences} n_{ij} means that frequencies of a subjective words ($w_i(1 \leq i \leq k)$) in a subjective and objective category ($c_j(j = 1, 2)$), its corresponding probability is $p_{ij} = n_{ij}/n$, n is sum of all the n_{ij} .

As shown in Eq. (1), the probability table is expressed as logarithmic form.

$$\eta_{ij} = \ln p_{ij} = \ln \left(\frac{p_{i\cdot} \cdot p_{\cdot j}}{p_{i\cdot} \cdot p_{\cdot j}} \right) = \ln p_{i\cdot} + \ln p_{\cdot j} + \ln \frac{p_{ij}}{p_{i\cdot} \cdot p_{\cdot j}}. \quad (1)$$

$$\text{Make } \eta_{i\cdot} = \sum_{j=1}^2 \eta_{ij}, \eta_{\cdot j} = \sum_{i=1}^k \eta_{ij}, \eta_{\cdot\cdot} = \sum_{i=1}^k \sum_{j=1}^2 \eta_{ij}.$$

So, the average logarithmic probability can be calculated by the following formula (2, 3, and 4).

$$\bar{\eta}_{i\cdot} = \frac{1}{2} \sum_{j=1}^2 \eta_{ij} \quad (2)$$

$$\bar{\eta}_{\cdot j} = \frac{1}{k} \sum_{i=1}^k \eta_{ij} \quad (3)$$

$$\bar{\eta}_{\cdot\cdot} = \frac{1}{2 \times k} \sum_{i=1}^k \sum_{j=1}^2 \eta_{ij}. \quad (4)$$

Make $\gamma_{ij} = \eta_{ij} - \bar{\eta}_{i\cdot} - \bar{\eta}_{\cdot j} + \bar{\eta}_{\cdot\cdot}$, $\hat{p}_{1j} = n_{1j}/n$, $\hat{p}_{i\cdot} = n_{i\cdot}/n$, $\hat{p}_{\cdot j} = n_{\cdot j}/n$.

And γ_{ij} is the interaction between words w_i and subjectivity Category C_j . $\gamma_{ij} > 0$ presents there is positive interaction, and $\gamma_{ij} < 0$ presents they have a reverse effect on the interaction, and when $\gamma_{ij} = 0$, there is no interaction between them. We also define $\hat{\eta}_{ij}$, $\hat{\eta}_{i\cdot}$, $\hat{\eta}_{\cdot j}$ and $\hat{\eta}_{\cdot\cdot}$ as follows:

$$\hat{\eta}_{ij} = \ln \hat{p}_{ij} = \ln n_{ij} - \ln n \quad (5)$$

$$\hat{\eta}_{i\cdot} = \ln \frac{1}{2} \sum_{j=1}^2 \eta_{ij} = \frac{1}{2} \sum_{j=1}^2 \left(\ln \frac{n_{ij}}{n} \right) = \frac{1}{2} \sum_{j=1}^2 (\ln n_{ij}) - \ln n \quad (6)$$

$$\hat{\eta}_{\cdot j} = \frac{1}{k} \sum_{i=1}^k \eta_{ij} = \frac{1}{k} \sum_{i=1}^k \left(\ln \frac{n_{ij}}{n} \right) = \frac{1}{k} \sum_{i=1}^k (\ln n_{ij}) - \ln n \quad (7)$$

$$\hat{\eta}_{\cdot\cdot} = \frac{1}{2k} \sum_{i=1}^k \sum_{j=1}^2 \eta_{ij} = \frac{1}{2k} \sum_{i=1}^k \sum_{j=1}^2 \left(\ln \frac{n_{ij}}{n} \right) = \frac{1}{2k} \sum_{i=1}^k \sum_{j=1}^2 (\ln n_{ij}) - \ln n. \quad (8)$$

And further we can calculate the estimated value of γ_{ij} by the Eq. (9)

$$\begin{aligned} \hat{\gamma}_{ij} &= \hat{\eta}_{ij} - \hat{\eta}_{i\cdot} - \hat{\eta}_{\cdot j} + \hat{\eta}_{\cdot\cdot} \\ &= \ln n_{ij} - \frac{1}{2} \sum_{j=1}^2 (\ln n_{ij}) - \frac{1}{k} \sum_{i=1}^k (\ln n_{ij}) + \frac{1}{2 \times k} \sum_{i=1}^k \sum_{j=1}^2 (\ln n_{ij}). \end{aligned} \quad (9)$$

We use $\hat{\gamma}_{ij}$ to measure the contribution of candidate words (w_i) to subjective category (C_j), $\hat{\gamma}_{ij}$ shows the words' subjectivity weight. Table 2 shows the value ($\hat{\gamma}_{ij}$) of candidate subjective words in the training corpus.

4 The Identification of Subjective Words in Fuzzy Sets

Depending on the weight of the subjectivity of the words, we will divide them into fuzzy sets, namely: 'high subjective intensity', 'moderate subjective intensity', 'low subjective intensity', then we construct the membership function of each collection, according to the membership function to determine the subjective intensity of the unknown words.

4.1 Membership Function of the Intensity of the Subjective Words

We selected trigonometric functions as membership function to describe the distribution. Firstly, make cluster centers of three-level collection $M = \{m_1, m_2, m_3\}$, and then we defined the membership function as follows:

$$T_l(x) = \begin{cases} 1 & x \leq m_1 \\ \frac{m_2 - x}{m_2 - m_1} & m_1 < x < m_2 \\ 0 & x \geq m_2 \end{cases} \quad (10)$$

Table 2 The weight of some opinion indicators and sentiment words under log-linear modeling

Category of subjective words	Examples	$\hat{\gamma}_{ij}$
Opinion indicator words	'Feel'	3.6310
	'Indicate'	3.6042
	'Assert'	2.3900
	'Forecast'	2.3233
	'Report'	-1.7848
	'Cassette'	-2.1916
Sentiment words	'Inevitable'	2.3510
	'Satisfied'	2.2740
	'Afraid'	2.0715
	'Pollution'	1.9279
	'Accept'	-0.5214
	'Issue'	-0.5389

$$T_{med}(x) = \begin{cases} 0 & x \leq m_1 \\ \frac{x-m_1}{m_2-m_1} & m_1 < x < m_2 \\ \frac{m_3-x}{m_3-m_2} & m_2 < x < m_3 \\ 0 & x \geq m_3 \end{cases} \quad (11)$$

$$T_h(x) = \begin{cases} 1 & x \geq m_3 \\ \frac{x-m_2}{m_3-m_2} & m_2 < x < m_3 \\ 0 & x \leq m_2 \end{cases} \quad (12)$$

In this paper, we use the method of self-organizing feature maps to determine the center collection M , the method of SOM has corrected the distance of sample point to the center point by the method of error propagation and via an iterative convergence ultimately determine the cluster center. According to the SOM algorithm we can calculate the weight set of cluster centers of opinion indicator words and sentiment words. Among the membership function of the views indicator words $m_1 = -1.226$, $m_2 = 1.035$, $m_3 = 3.890$, in the membership function of the sentiment words $m_1 = -0.854$, $m_2 = 1.205$, $m_3 = 3.114$ [6].

4.2 Subjective and Objective Classification Based on Complex Rules

To test the impact of the words subjective intensity on Chinese sentence subjectivity classification, we use a set of classifier based on rules, which mainly determined sentence's subjectivity by looking for the different subjective intensity of the subjectivity of the words in the sentences. Unlike (Riloff and Wiebe 2003) [7] using the single rule classifier, it mainly contains the following rules.

- If the sentence contains lots of high intensity or medium intensity view indicator words whose intensity is greater than a given threshold value δ , then sentence is subjectivity sentence.
- If the sentence contains lots of high intensity or medium intensity sentiment words whose intensity is greater than a given threshold value δ , then sentence is subjectivity sentence.
- If the sentence contains lots of high-intensity or medium intensity view indicator words whose intensity is greater than a given threshold value δ , then sentence is subjectivity sentence; at the same time, the intensity of high intensity or medium intensity sentiment words contained in the sentence is greater than a given threshold value μ , then the sentence is subjectivity sentence.

δ and μ are two experienced threshold value which can be determined by experiment. As it can be seen from the above rules, rule 1 and rule 2 are view indicator words' and sentiment words' single effect on the subjective identification of sentence, rules 3 taking into account both.

5 Experiment and Analysis

5.1 Experiment Setting

Data we used in this paper are from the E-Learning platform academic text. We constructed our own training sets and test sets by extracting the data from these texts and the evaluation criteria we use is Lenient-AWK: recall rate (R), precision rate (P) and their harmonic mean (F) [8].

5.2 Experiment Result

5.2.1 Effect of View Indicator Words and Sentiment Words on Sentence Subjectivity Identification

The first sets of experiments were to test the effect of different types of subjective words on the recognition of subjectivity of Chinese sentences, including view indicator and sentiment words. The experiments in this article use the following three types of rules to evaluate the impact of words in the sentence subjectivity classification, the experimental results are given in Table 3.

As can be seen from Table 4, a system was constructed according to the rules 1 and rule 2 to obtain a higher precision, but the recall rate is lower. Rule 3 has achieved the best overall performance, it embodies an active role in considering the view indicator words and sentiment words in Chinese Sentences subjectivity

Table 3 Basic statistics of the experimental data

Text type	Training data	Test data
Theme	35	14
Document	901	188
Sentence	13416	4655

Table 4 Evaluation results for different classifier using different rules with $\delta = 1$ and $\mu = 1$

Subjectivity and objectivity classifier	<i>P</i>	<i>R</i>	<i>F</i>
Rule 1	0.7085	0.5551	0.6225
Rule 2	0.8144	0.5363	0.6447
Rule 3	0.7175	0.8006	0.7567

recognition. Rule 2 creates the classifier that has obtained the highest accuracy rate to 81.44 %, but the recall rate of only 53.63 %, the main reason may be the opinion sentences containing sentiment words in text, but the subjectivity of these sentiment words' intensity is weak, so the classifier constructed in rule 2 does not recognize too many subjective sentences.

5.2.2 Distinguish the Effect of Subjective Words' Intensity on Subjective and Objective Recognition of the Sentences

The second set of experiments was designed to test the subjectivity intensity of the words in the subjective and objective classification of the sentence [9]. We compared the experimental results with the Baseline system in the NTCIR-7MOAT and with the best WIA-Opinmine system. Our sentiment dictionary contains 8,596 subjectivity words, mainly from the CUHK and NTU sentiment dictionary. In this experiment, the Baseline system did not distinguish sentence subjectivity intensity, namely, the subjectivity of the level of intensity of the sentence is completely ignored, the WIA-Opinmine with a fine-grained to coarse-grained strategy to explore the composite characteristics sentence, and in the recognition of the subjectivity of the sentence, and ultimately achieve very good results.

Table 5 shows the results of the second set of experiments. We found that this system's precision and F values are higher than the WIA-Opinmine system, but lower in recall rate. Perhaps we use less subjective characteristics than WIA-Opinmine system, and therefore cannot recognize more subjective sentences in corpus. Our system is beyond the Baseline system about 10 %, which indicates that distinguish the subjectivity of the words in the intensity of subjective identification of the sentence has a very important role.

In order to study the key role of words' subjective intensity in subjectivity identification of Chinese sentence, the paper improved a subjective intensity learning method based on the log-linear model and fuzzy sets. Including:

Table 5 Comparison of our system with the best system at NTCIR-7 under the lenient standard

System	<i>P</i>	<i>R</i>	<i>F</i>
Baseline system	0.5288	0.8511	0.6523
WIA-Opinmine system	0.6520	0.8698	0.7453
Our system	0.7175	0.8006	0.7567

(1) candidate subjective words extraction and weighting terms; (2) constructed and parameters determined of membership function of subjective words with different intensity; (3) methods with the subjective and objective classification of sentences based complex rules. Experimental results show that the view indicator Verbs and sentiment words play a very important role in the classification of subjective sentences, and each of them are in a different way to boot opinions expressed in the sentence.

Even though the distinction the subjective intensity of words in the sentence can make subjective and objective classification results significantly improved, there is still insufficient during rough set reduction, for those with subjective and objective sentence model, whether to retain or reduction will affect the system in the implementation of judgment subjective or objective of recall and precision. Therefore, our follow-up work will improve it.

Acknowledgment First of all, I would like to thank my mentor the Professor Wang. Without his help I could not complete it successfully. Secondly, I sincerely thank the support and encouragement that my classmates give me. Thirdly, I would like to thank the support of the National Natural Science Foundation of China, and Beijing Natural Science Foundation. And finally I would show special thanks to my family for their support understanding and encouragement.

References

1. Qing Y, Zi-qiong Z, Zhen-xiong L (2007) Study of methods of subjectivity automatically discriminating of Chinese on sentiment analysis in Internet comments. *China J Inf Syst* 1(1):79–91
2. Ellen R, Siddhart P, Janyee W (2006) Feature subsumptionf or opinion analysis. In: *Proceedings of EMNLP'06, Sydney, Australia*, pp 440–448
3. Xin-fan M, Hou-feng W (2009) Research about effect of the context of factors on Subjective and objective recognition. *China Acad J*, 594–599
4. Long-shu L, Xiao-hong Z, Zhi-wei Z (2011) The subjective and objectivity Research of the Chinese text based on rough set theory. *Comput Technol Dev*, 114–115
5. Bo Z (2011) Chinese views sentence extraction based on SVM. The Academy of Computer of Beijing University of Posts and Telecommunications, Beijing
6. Xi W (2011) Research of methods on multi-granularity fusion of Chinese sentences' subjectivity and sentiment classification. The Academy of Computer Science of Heilongjiang University, Harbin
7. Ellen R, Janyce W, Phillips W (2003) Learning subjective nouns using extraction pattern bootstrapping. In: *Proceedings of CoNLL'03:25-3*

8. Yohei S, David KE, Lun-Wei K, Hsin-His C, Noriko K, Chin-Yew L (2007) Overview of opinion analysis pilot task at NTCIR-6. In: Proceedings of NTCIR-6 workshop meeting, Tokyo, Japan, pp 265–278
9. Yohei S, David KE, Lun-Wei K, Le S, Hsin-His C, Noriko K (2008) Overview of multilingual opinion analysis task at NTCIR-7. In: Proceedings of NTCIR-7 workshop meeting, Tokyo, Japan, pp 185–203



<http://www.springer.com/978-3-642-54929-8>

Knowledge Engineering and Management
Proceedings of the Eighth International Conference on
Intelligent Systems and Knowledge Engineering,
Shenzhen, China, Nov 2013 (ISKE 2013)
Wen, Z.; Li, T. (Eds.)
2014, X, 406 p. 123 illus., Softcover
ISBN: 978-3-642-54929-8