

Rule-Based Weibo Messages Sentiment

Polarity Classification towards Given Topics

Hongzhao Zhou,
School of Chinese
Language and Literature,
Communication University of
China, Beijing, China
zhzwin2011@163.com

Yonglin Teng, Min Hou, Wei He,
National Broadcast Media Language Resources
Monitoring & Research Center,
Communication University of China, Beijing,
China
{tengyonglin, houmin, hewei}@cuc.edu.cn

Hongtao Zhu, Xiaolin Zhu and Yanfei Mu
School of Chinese Language and Literature,
Communication University of China, Beijing,
China
{727134880, 893136856, 610202467}@qq.com

Abstract

Weibo messages sentiment polarity classification towards given topics refers to that the machine automatically classifies whether the weibo message is of positive, negative, or neutral sentiment towards the given topic. The algorithm the sentiment analysis system CUCsas adopts to perform this task includes three steps: (1) whether there is an “exp” (short for “expression having evaluation meaning”) in the weibo message; (2) whether there is a semantic orientation relationship between the exp and topic; (3) the sentiment polarity classification of the exp. CUCsas completes step (1) based on the sentiment lexicon and sentiment value assignment rules, completes step (2) based on the topic extraction and sentiment polarity classification rule base, and completes step (3) based on the sentiment computing rules. Taking 20 given topics and a total of 19,469 weibo messages released by SIGHAN-2015 Bake-off as the test data, the overall F value of the rule-based system CUCsas is 0.69 in the unrestricted test.

1 Algorithm Description

The locutionary subjectivity denotes the locutionary agent’s self-expression of cognition, feeling or perception in the use of language (John Lyons, 1995). And the evaluation is one type of locutionary subjectivity. An evaluation discourse con-

sists of four basic elements: $E(s) = \{\text{sub}, \text{obj}, \text{exp}, \text{com}\}$. Herein, “E(s)” represents an evaluation discourse, and “sub”, “obj”, “exp” and “com” refers to the subject of evaluation, the object of evaluation, an expression having evaluation meaning, and other discourse components respectively (Zhou Hongzhao et al., 2014). The study of this paper is under the condition of knowing obj (= the given topic) in the weibo message, enabling the system automatically recognize whether there is an exp in the same weibo message. If there is not, the system will output result [topic: 0]; if there is, the system will make a further identification that whether there is a semantic orientation relationship between the exp and the given topic. If there is not, the system will outputs result [topic 0]; if there is, the system will further classify the sentiment polarity of the exp. If it is positive, the system will output result [topic 1]; if it is negative, the system will output result [topic -1]; if it is neutral, the system will output result [topic 0]. Apparently, the algorithm is different from some widely used machine learning sentiment polarity classification algorithms, such as Naïve Bayes, Max Entropy, Boosted Trees and Random Forest (Amit Gupte et al., 2014). Figure 1 shows the algorithm of the the system of rule-based weibo messages sentiment polarity classification towards given topics.

Example (1) <weibo>:三星发布 Galaxy S6 和 S6 Edge, 下月正式开卖。 </weibo> (There is no exp in the weibo message. → Output: 0)

Example (2) <weibo>:评论员手好丑,评论的也很垃圾,不看了//【视频:三星 GALAXY S6 初体验】</weibo> (There are exps “好丑” and “垃圾” in the weibo message. → But there is no semantic orientation relationship between the exps and the given topic “三星 S6”. → Output: 0)

Example (3) <weibo>:三星 s6 奇丑无比,边框还仿苹果.</weibo> (There is an exp “奇丑无比” in the weibo message. → There is a semantic orientation relationship between the exp and the given topic “三星 S6”. → The sentiment polarity of the exp is negative. → Output: -1)

Example (4) <weibo>:HTC One M9 与三星的 S6 哪个更惊艳?</weibo> (There is an exp “惊艳” in the weibo message. → There is a semantic orientation relationship between the exp and the given topic “三星 S6”. → The sentiment polarity of the exp is neutral in the weibo message context. → Output: 0)

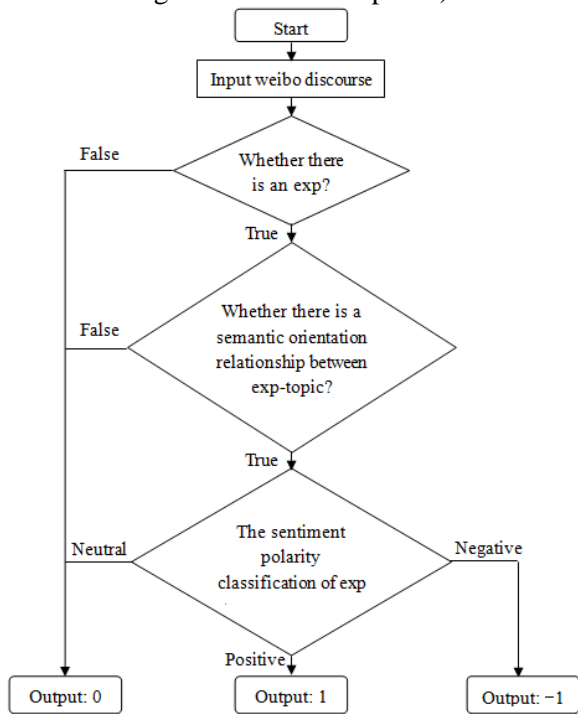


Figure 1. The Algorithm of the Weibo Topic Sentiment Polarity Classification

2 The Automatic Recognition of the Exp in the Weibo Message

From the perspective of linguistics, the exp can be divided into three broad categories, including six specific types.

(1) Category one

<1> Type one: the context-free evaluation word or phrase

Feature: Independent of context, it expresses positive or negative evaluation meaning by itself.

Sentiment marker: po or ne

Examples: 漂亮、败类、狗仗人势

Total in the sentiment lexicon: 26,042

(2) Category two: the context-sensitive evaluation word or phrase

Feature: Whether it expresses evaluation meaning or not depends on the context.

<2> Type two: the commendatory potential word

Feature: When modified by the degree word, it can express positive evaluation meaning.

Semantic marker: pxn

Examples: 规范、人道、man

Total in the semantic lexicon: 51

<3> Type three: the derogatory potential word

Feature: When modified by the degree word, it can express negative evaluation meaning.

Semantic marker: nxn

Examples: 封建、一般、2

Total in the semantic lexicon: 18

<4> Type four: the meaning-shifting noun

Feature: When modified by the affirmative word such as 有 or 具有, it expresses positive evaluation meaning; when modified by the negative word such as 没有 or 毫无, it expresses negative evaluation meaning.

Semantic marker: ypn

Examples: 诚信、效率、素质

Total in the semantic lexicon: 198

<5> Type five: the adjective of weights and measures

Feature: When combined with the product attribute or human character word, the adjective of weights and measures, such as 高、低、大、小, can express evaluation meaning.

Examples: 清晰度+高、油耗+低、辐射+大

Total in the phrase rule base: 153

(3) Category three

<6> Type six: Evaluation syntactical structure or distant collocation.

Examples: 无法和……相比; 引发……问题

Total in the phrase rule base: 52

2.1 The Storage and Formal Description of Different Types of Exps

(1) Words and phrases of type one are stored in the sentiment lexicon SentiDic.txt in the form of entries. The lexicon format and entry samples are as follows:

[Word or phrase sentiment intensity value sentiment intensity value]	Part of speech	Positive Negative	Positive sentiment intensity value]
漂亮	a	0.5	0
鄙视	v	0	0.5
败类	n	0	0.5

(2) Words and phrases of type two, three and four are stored in the semantic dictionary *Usr-Di1.dic* first. Then, corresponding sentiment value assignment rules for them are formulated in the phrase rule base *PhraseRule.txt*.

The lexicon format and entry samples:

[Word or phrase	Semantic marker]
规范	pxn
封建	nxn
诚信	ypn

The sentiment value assignment rule samples:

① $*/mopo + */pxn = \#2:0.75$

The left part of = is the matching condition, the right part of = is the operation result. The symbol $*/mopo$ represents a degree modifier (e.g. 很、非常、十分). The function of this rule: When there is a $*/mopo$ in front of $*/pxn$, a 0.75 sentiment value is assigned to $*/pxn$.

② $*/mone + */pxn = \#2:-0.5$

The symbol $*/mone$ represents a negative modifier (e.g. 没有、毫无、缺乏). The function of this rule: When there is a $*/mone$ in front of $*/pxn$, a -0.5 sentiment value is assigned to $*/pxn$.

(3) As to type five and six, corresponding sentiment value assignment rules are formulated in the phrase rule base *PhraseRule.txt*. The sentiment value assignment rule samples:

③ 质量|性能|像素|分辨率|清晰度|安全系数
/% + #[*!/(w|mone)] + 高/a = #3:0.5

The symbol $\#[*!/(w|mone)]$ means that the rule can cross arbitrary segmentations here except the punctuation(w) or negative modifier(mone).

Example (5) <weibo>:丰田车的安全系数的确是低了点。</weibo> (It satisfies the matching condition of rule ③, so a 0.5 sentiment value is assigned to the third item 低/a.)

Module 1	the exclusive method
Explanation	When the evaluation object of the exp is non-topic, the system will assign a 0 sentiment value to the topic, so as to avoid the weibo message continuing to match the latter rule modules and cause errors.
Rule sample	QSB + #[*!/(w topic)] + */(NP)&!(topic v1) + #[*!/(w topic)] + 是/% + #[*!/(w)] + *//topic + #[*!/(w)] + *//v1&(n in ln) + #[*!/(w)] + *//w y e \$ = N7:0
Rule sample explanation	(1) QSB: It is a macro definition symbol (including the punctuation, conjunction, evaluation-triggering word, time word or discourse maker) used as the initial

④ 无法|没法|不能|不可能/v + 和|跟|同|与/p + #[*!/(w)] + 比|相比/% = #1:-0.5

Example (6) <weibo>:三星 S6 的屏幕分辨率根本无法和 iPhone6 相比。</weibo> (It satisfies the matching condition of rule ④, so a -0.5 sentiment value is assigned to the first item 无法/v.)

Based on the sentiment lexicon *SentiDic* and sentiment value assignment rules in *PhraseRule*, the system *CUCsas* realizes the automatic recognition of whether there is an exp in the weibo discourse. Figure 2 shows the recognition procedure:

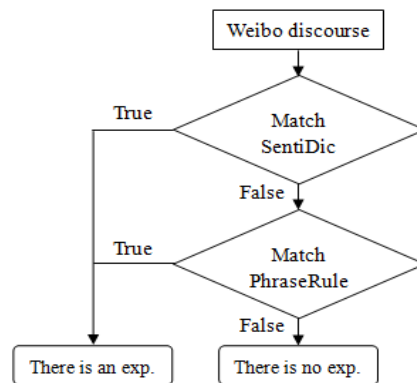


Figure 2. The Procedure of the Exp Recognition

3 The Identification of Whether There Is a Semantic Orientation Relationship between Exp-Topic

The existence of an exp in the weibo message does not imply a semantic orientation relationship between the exp and the topic. Because the evaluation object of the exp has two potential choices: topic or non-topic. The system *CUCsas* uses the method of combining syntactic structure and semantic features to build a topic extraction and polarity classification rule base. The essence of the rule base is using formal languages to describe the definite semantic direction relationships between exp-topic, which are induced by analyzing the training corpus by us. The topic extraction and polarity classification rule base consists of 10 rule modules with a total of 36 rules (see Table 1).

	item in this rule; (2) NP: It is a macro definition symbol (including the common noun or proper noun such as the name of a person, organization or product) representing a nominal element; (3) */topic: the given weibo topic; (4) */vl: an exp.
Matching example	Topic:雾霾 <weibo>:原来我一直以为汽车尾气排放是 <u>雾霾</u> 的罪魁祸首。 </weibo> [output: 雾霾 0]
Rule number	1-7
Module 2	the adversative compound sentence
Explanation	The content behind the adversative word is usually the semantic focus which the speaker wants to convey. Hence the rule only selects the exp appearing after the adversative word and semantically oriented to the topic as the output result, ignoring the other exps in the weibo message.
Rule sample	QSB + #[*/*%] + */topic + #[(; ; . .)!NP] + ZZC/% + #[, ,/%] + #[(, ; ; . . ? ! 、)!NP xjc] + */vl&!hzv + #[(?!? 吗 呢 么)!xjc] + JSB = N3:N8
Rule sample explanation	(1) ZZC: It is a macro definition symbol (including a total of 23 adversative words, such as 但、但是、可是、而是、然而、反而、却); (2) =N3:N8: It means assigning the sentiment value of the eighth item */vl&!hzv to the third item */topic.
Matching example	Topic:三星 S6 <weibo>:本以为三星快不行的时候, <u>S6</u> 却 <u>震撼</u> 登场了。 </weibo> [output: S6 1]
Rule number	8-10
Module 3	topic-exp co-occurrence in the same clause
Explanation	When the topic and the exp appear in the same clause, the rule will select the exp nearest to the topic as the one semantically oriented it.(The exception is that the topic is the subject of a sentence expressing a causing or obtaining meaning or with a “preposition + object” adverbial.) In addition, according to the Chinese pragmatic habit that the semantic focus is usually located at the end of the discourse, when exps appear both before and after the topic, i.e. exp1-topic-exp2, the rule will select exp2 only as the output result.
Rule sample	QSB + #[(比 把)!xjc] + */topic + #[*!w xjc vl nq] + */vl&!hzv + #[(?!? 吗 呢 么)!xjc] + JSB = N3:N5
Rule sample explanation	*/vl&!hzv: The exp is arbitrary except for the backward-orientated sentiment verb(hzv) such as 喜爱、佩服 or 鄙视, because the evaluation object of the hzv is usually the component after it, not the topic before it.
Matching example	Topic:雾霾 <weibo>:我赞成中国 <u>雾霾</u> 问题非常 <u>严重</u> 。 </weibo> [output: 雾霾 -1]
Rule number	11-17
Module 4	the sentence expressing a causing or obtaining meaning
Explanation	When the topic is the subject of a sentence expressing a causing or obtaining meaning, the rule will select the last exp in the clause introduced by a word expressing a causing or obtaining meaning as the output result.
Rule sample	QSB + #[*!vl xjc] + */topic + #[!. /!NP xjc] + TSC/% + #[*!w topic xjc] + */vl + #[(?!? 吗 呢 么)!xjc] + JSB = N3:N7
Rule sample explanation	TSC: It is a macro definition symbol (including a total of 31 words expressing a causing or obtaining meaning, such as 让、使得、引起、导致、成为 or 得到).
Matching example	Topic:中国人疯抢日本马桶 <weibo>:其中最为热销的产品竟然是智能 <u>马桶盖</u> , 卖到几近断货, 真是让人大跌眼镜。 </weibo> [output: 马桶盖 -1]
Rule number	18
Module 5	the sentence with a “preposition + object” adverbial
Explanation	When the topic is the subject of a sentence with a “preposition + object” adverbial, the rule will select the exp in the central components modified by the adverbial as the output result.

Rule sample	QSB + #[*!/vl xjc] + */topic + #[!。 /!NP xjc] + 对 对于 为 将 给/p + #[!(; ; 。 ,? ! ;)/!topic xjc] + */vl&!(hzv xlv) + #[!(?!? 吗 呢 么)/!xjc] + JSB = N3:N7
Rule sample explanation	*/vl&!(hzv xlv): The exp is arbitrary except for the backward-orientated sentiment verb(hzv) or psychological sentiment verb(xlv), because the evaluation object of the hzv or xlv is usually the object of the preposition, not the topic as the subject of the sentence.
Matching example	Topic:央行降息 <weibo>:羊年第一个周末央行再度出手 <u>降息</u> , 对券商、保险、地产等绝大多数品种构成较大利好。 </weibo> [output: 降息 1]
Rule number	19
Module 6	the comparative sentence
Explanation	When the topic serves as the comparative subject in the comparative sentence, its sentiment vale = the sentiment value of the exp serving as the comparative result; when the topic serves as the comparative datum in the comparative sentence, its sentiment vale = the sentiment value of the exp serving as the comparative result $\times (-1)$ (Zhou Hongzhao et al., 2014).
Rule sample	QSB + #[*!/vl xjc] + */topic + #[(。 ; ?)/!vl xjc] + 比 相比 比起 对比/p + #[(。 ! ? ;)/!topic xjc] + */vl + #[(?!? 吗 呢 么)/!xjc v] + JSB = N3:N7
Rule sample explanation	The */topic (N3) is located before the comparative-marker word 比 相比 比起 对比/p(N5) .So it serves as the comparative subject and its sentiment vale = the sentiment value of the exp */vl(N7) serving as the comparative result.
Matching example	Topic:三星 S6 <weibo>:个人感觉 <u>S6</u> 前面板一如既往三星风格, 背面更是比 iPhone6 还难看。 </weibo> [output: S6 -1]
Rule number	20-24
Module 7	the causation compound sentence
Explanation	In the causation compound sentence, the exp may appear in the reason clause, while its evaluation object appears in the result clause.
Rule sample	*/topic + #[(。 ? ! ; ;)/!xjc] + 因为/% + #[*!/w] + */vl = N1:N5
Rule sample explanation	In module 4, the topic is the reason, while the exp is the result. Here, the topic is the result, while the exp is the reason. The two rule modules complement each other.
Matching example	Topic:中国人疯抢日本马桶 <weibo>:终于明白为什么中国人都要去日本买 <u>马桶盖</u> 了, 因为好用到飙泪! </weibo> [output: 马桶盖 1]
Rule number	25
Module 8	The topic and the exp are distributed in different clauses or sentences. Type one: topic + exp
Explanation	The topic appears first, and then the exp appears in the clause or sentence adjacent or nonadjacent to the clause or sentence the topic in. In this case, only the weibo message satisfies certain syntactic and semantic constraints, will the rule judge that the evaluation object of the exp is the topic.
Rule sample	QSB + #[*!/vl xjc] + */topic + #[!。 /!vl xjc] + */w + #[!。 /!xjc NP] + */vl + #1:3[!(吗 呢 么)/u y e] + JSB = N3:N7
Rule sample explanation	Constraints of the rule sample: (1) There is no exp appearing together with the topic in the clause; (2) There is no NP appearing before the exp in the clause; (3) The word class after the exp is only auxiliary, modal or interjection, and three interrogative words 吗、呢 and 么 are forbidden.
Matching example	Topic:油价 <weibo>:涨 <u>油价</u> 的时候也不提消费税了, <u>流氓</u> 啊 </weibo> [output: 油价 -1]
Rule number	26-32
Module 9	The topic and the exp are distributed in different clauses or sentences. Type two: exp + topic

Explanation	The exp appears first, and then the topic appears in the clause or sentence adjacent or nonadjacent to the clause or sentence the exp in. In this situation, only the weibo message satisfies certain syntactic and semantic constraints, will the rule judge that the evaluation object of the exp is the topic.
Rule sample	*/^ + #[*!/nq] + */na + #[*!/w] + */vl + #[*!/nq] + */topic&nq = N7:N5
Rule sample explanation	Constraints of the rule sample: (1) */^: The initial item of the rule is the weibo start marker; (2) #[*!/nq]: The word with a semantic marker of product name is forbidden; (3) */na: A word with the semantic marker of product attribute must appear; (4) */topic&nq: The topic word must be also a product name.
Matching example	Topic:三星 S6 <weibo>:电池是唯一的小遗憾//【沉默后的爆发 三星 Galaxy S6 竞争力分析】 http://t.cn/RwQ6plU (分享自 @鲜果) </weibo> [output: S6 -1]
Rule number	33-35
Module 10	anaphora resolution
Explanation	When the referent of a pronoun is the topic, the rule will assign the sentiment value of the exp semantically orientated to the pronoun to the topic.
Rule sample	*/topic + #[*!/xjc vl NP] + 你 你们 这 这些 这样 这么 此举/r + #[*!/m q] + #[*!/w xjc vl] + */vl + #[*!/(! ?) 吗 呢 么)/!nr xjc] + */\$ = N1:N6
Rule sample explanation	(1) #[*!/m q]: a numeral or quantifier can appear or not appear here; (2) */\$: the end marker of the weibo message.
Matching example	Topic:油价 <weibo>:在未来一两年我们会看到国际油价的触底。这种状况会很好的帮助中国、日本开辟出新的机遇。</weibo> [output: 油价 1]
Rule number	36
Note:	(1) The 36 rules of the 10 rule modules are sequentially arranged, forming the topic extraction and sentiment polarity classification rule base. (2) Matching procedure: The weibo message matches the rule base starting from the first rule. If the matching succeeds, the system will output a corresponding matching result; if fails, the weibo message will skip to the second rule to continue matching. If this matching succeeds, the system will output a corresponding matching result; or else the weibo message will skip to the next rule to continue matching.....If the matching still fails at the end of the rule base (i.e. rule 36), then the system will make a judgment that there is no semantic orientation relationship between the exp and the topic in this weibo message and output a corresponding result: topic 0. The next weibo message matches the rule base in the same way.....until the last weibo message in the experimental data.

Table 1. Topic Extraction and Sentiment Polarity Classification Rule Base

Based on the topic extraction and polarity classification rule base, the system CUCsas realizes the automatic identification of whether there is a semantic orientation relationship between the exp and the topic in the weibo message. If the weibo message matches the rule base unsuccessfully, the system will output topic 0; if successfully, the system will assign the value of the corresponding exp to the topic. If the value > 0, the system will output: topic 1; if the value < 0, the system will output: topic -1; if the value = 0, the system will output: topic 0. Figure 3 shows the general procedure:

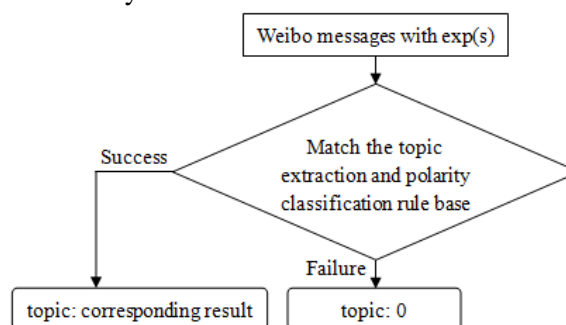


Figure 3. The Procedure of Topic Extraction and Sentiment Polarity Classification

4 The Sentiment Polarity Classification of the Exp

The term “corresponding result” in Figure 3 contains double meanings: i The “corresponding”

means that there is a semantic orientation relationship between the exp and the topic. ii The “result” refers to the sentiment value and polarity of the exp in the weibo message context, not necessarily equals the value and polarity in the sentiment lexicon. i is guaranteed by 36 rules of 10 modules. ii is obtained by sentiment computing rules (see Table 2) in the PhraseRule.txt.

Type 1	Contrary
Description	The sentiment polarity of the exp in the weibo message context is contrary to its sentiment polarity in the sentiment lexicon.
Features	(1) The exp is modified by the word with a negative semantic marker “mone”; (2) The exp appears in a negative sentence pattern characterized by words such as 难道 or 怎么可能; (3) The exp appears in the special collocation characterized by specific words. For instance, the sentiment polarity of 美化 is positive in the sentiment lexicon, but when it collocates with 战争、侵略 or 历史, its sentiment polarity will turn negative.
Rule sample	*/mone + */po ne = N2*N1
Matching example	(三星 S6) (看样子) (一点) (都) ([不]好用:-1) (。)
Rules total	51
Type 2	Dissolution
Description	The evaluation meaning of the exp is dissolved in the weibo message context.
Features	(1) The exp appears in the sentence introduced by the word with an evaluation dissolving marker “xjc” such as 如果、假如、祝愿、但愿、能否、是否—30 in all; (2) The exp appears in an evaluation dissolving sentence pattern characterized by the collocation of specific words or word classes, such as 是...还是..., exp + vv.
Rule sample	*/xjc + #[*!/w] + */po ne = #3:0

Matching example	(三星 S6) (能否) (力挽狂澜:0) (?)
Rules total	12
Type 3	Consistency
Description	The polarity of the exp in the weibo message context is consistent with the sentiment lexicon. But the sentiment intensity can be unchanged, enhanced or weakened.
Features	(1) Features mentioned in type 1 and type 2 must not appear; (2) Features maintaining, enhancing or weakening the sentiment intensity of the exp, such as semantic markers or specific words can appear.
Rule sample	*/mopo + */po ne = N2*(1+N1)
Matching example	(三星 S6) (,) (外观) (确实) ([很]漂亮:0.875) (。)
Rules total	10

Table 2. Three Types of the Exp and

Corresponding Sentiment Computing Rules

Based on the sentiment computing rules stored in the PhraseRule, the system realizes the calculation of the sentiment value of the exp in the weibo message context.

5 Experimental Results and Analysis

Taking 20 given topics and a total of 19,469 weibo messages released by SIGHAN-2015 Bake-off as the test data, the experimental results of the sentiment analysis system CUCsas are as follows:

SIGHAN-2015 Bake-off (unrestricted test)	Precision	0.6937182
	Recall	0.6937182
	F	0.6937182
	Precision+	0.1839539
	Recall+	0.36024305
	F+	0.24354461
	Precision-	0.5010653
	Recall-	0.3877439
F-	0.4371805	

Table 3. The SIGHAN-2015 Bake-off (Unrestricted Test) Evaluation Result of CUCsas

Only using the sentiment lexicon resource, the experimental results are as follows:

SIGHAN-	Precision	0.46001335
---------	-----------	------------

2015 Bake-off (unrestricted test)	Recall	0.46001335
	F	0.46001335
	Precision+	0.12713068
	Recall+	0.62152778
	F+	0.2110849
	Precision-	0.34455307
	Recall-	0.6779335
	F-	0.45689415

Table 4. Only Using the Sentiment Lexicon

Using the sentiment lexicon together with the phrase rule base resource, the experimental results are as follows:

SIGHAN- 2015 Bake-off (unrestricted test)	Precision	0.48019929
	Recall	0.48019929
	F	0.48019929
	Precision+	0.13504006
	Recall+	0.59982639
	F+	0.22044983
	Precision-	0.34286523
	Recall-	0.66556746
	F-	0.45258339

Table 5. Using the Sentiment Lexicon
Together with the Phrase Rule Base

Comparing Table 4 with Table 5, we can see the introduction of the phrase rule base improved the system overall performance, but only to a small extent. Comparing Table 5 with Table 3, we can see the introduction of the topic extraction and polarity classification rule base further improved the system overall performance to a large extent.

At present, the overall F value of the system is about 0.69. Evaluation results in Table 3 suggest that the performance of the system is good in dealing with neutral sentiment weibo messages, but poor in dealing with positive sentiment weibo messages ($F+ \approx 0.24$) and negative sentiment weibo messages ($F- \approx 0.44$).

Reasons and solving methods for poor Recall+ and Recall- : (1) The scale of the topic extraction and polarity classification rule base built according to the training data is small (only 36 rules). Thus, the language phenomena having not appeared in the training data can't be covered. For instance, the module 10 — anaphora resolution neglects the case that the pronoun appears ahead of the topic. In the next stage, new rules will be added to the rule base to expand its coverage. (2) The sentiment lexicon and the sentiment phrase

rule base are not incomplete so that many exps in the test data can't be recognized. In the next stage, the system will improve the automatic recognition of unlisted exps.

Reasons and solving methods for poor Precision+ and Precision-: (1) Some rules in the topic extraction and polarity classification rule base do not appropriately describe the semantically orientated relationship between topic-exp, which leads to the wrong extraction of exps. In the next stage, some rules will be revised based on the errors analysis. (2) Some "exps" in the sentiment lexicon actually do not have evaluation meaning. For example, the word 激烈 is not a sentiment word. However, it is listed in the sentiment lexicon as a negative word. Therefore, the sentiment polarity output result of Topic :水货客 in <weibo>:反水货客行动越趋激烈。 </weibo> is wrong -1. In the next stage, the sentiment lexicon will be checked and non-sentiment words will be removed.

6 Conclusion

In this paper, firstly, we proposed the algorithm of rule-based weibo messages sentiment polarity classification towards given topics. Then, we adopted the rule methods to implement the requirements of the algorithm procedures. Based on the sentiment lexicon SentiDic and sentiment value assignment rules in PhraseRule, the sentiment analysis system CUCsas realized the automatic recognition of the exp in weibo messages. Based on the topic extraction and polarity classification rule base, the system realized the automatic identification of whether there is a semantic orientation relationship between the exp and the topic. And based on the sentiment computing rules in PhraseRule, the system realized the sentiment value calculation and polarity classification of the exp in specific weibo message context. At present, the overall F value of the ruled-based sentiment analysis system CUCsas is about 0.69. In the future, the lexicon and rule base will be revised based on the errors analysis to improve the performance of the system.

Reference

Amit Gupte, Sourabh Joshi, Pratik Gadgul, Akshay Kadam. 2014. Comparative Study of Classification Algorithms used in Sentiment Analysis. *International Journal of Computer Science and Information Technologies (IJCSIT)*, Vol. 5 (5):6261-6264.

John Lyons. 1995. *Linguistic Semantics: An Introduction*. Cambridge University Press, Cambridge. UK.

ZHOU Hongzhao, HOU Mingwu, YAN Pengli, ZHANG Yeqing, HOU Min and TENG Yonglin. 2014. Function of Semantic Features in Opinion Target Extraction and Its Polarity Identification. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 50(1):93-99.

ZHOU Hongzhao, HOU Mingwu, HOU Min and TENG Yonglin. 2014. Chinese Comparative Sentences Identification and Comparative Elements Extraction Based on Semantic Classification. *Journal of Chinese Information Processing*, 28(3):136-141.