International Journal of Advanced Intelligence
Volume 2, Number 1, pp.105-117, July, 2010.

Sentence Emotion Analysis and Recognition Based on Emotion Words Using Ren-CECps*

Changqin Quan
Faculty of Engineering, University of Tokushima
2-1 Minamijosanjima, Tokushima 770-8506, Japan
quan-c@is.tokushima-u.ac.jp

Fuji Ren
Faculty of Engineering, University of Tokushima
2-1 Minamijosanjima, Tokushima 770-8506, Japan
ren@is.tokushima-u.ac.jp

Received (January 2010)
Revised (May 2010)

Emotion recognition on text has wide applications. In this study, we make an analysis on sentence emotion based on emotion words using Ren-CECps (a Chinese emotion corpus). Some classification methods (including C4.5 decision tree, SVM, NaiveBayes, ZEROR, and DecisionTable) have been compared. Then a supervised machine learning method (Polynomial kernel method) is proposed to recognize the eight basic emotions (Expect, Joy, Love, Surprise, Anxiety, Sorrow, Angry and Hate). Using Ren-CECps, we get the emotion lexicons for the eight basic emotions. Polynomial kernel (PK) method is used to compute the similarities between sentences and the eight emotion lexicons. Then the experiential knowledge derived from Ren-CECps is used to recognize whether the eight emotion categories are present in a sentence. The experiments showed promising results.

Keywords: Emotion recognition; Chinese emotion corpus; Classification methods; Polynomial kernel.

1. Introduction

Emotions play important role in human intelligence, rational decision making, social interaction, perception, memory, learning, creativity, and more [1]. Given the importance of computer mediated communication (CMC), a lot of researches have been done on emotion recognition [2,3]. At the same time, textual emotion recognition is increasingly attracting attention.

Previous researches for emotional analysis of texts have included a variety of text contents: weblogs[4,5], stories [6,7], news[8], text messages [9], spoken dialogs[10], etc. For many applications, identifying emotions only on document level may not be sufficient. A text-based emotion prediction system would benefit from identifying the

*The prior version of the paper has been presented in international conference NLP-KE 2009.

emotional affinity of sentences. The emotion analysis on sentence level may also be important for more detailed emotion analysis systems. Alm explored the text-based emotion prediction problem. In order to classify the emotional affinity of sentences in the narrative domain of children's fairy tales, they annotated a corpus of 22 Grimms' tales on sentence level with eight emotion categories (Angry, disgusted, fearful, happy, sad, positively Surprised, negatively Surprised)[7]. Neviarouskaya addresseds the tasks of recognition and interpretation of affect communicated through text messaging[9]. Aman and Szpakowicz classify emotional and non-emotional sentences based on a knowledge-based approach[11]. Classifying the mood of a single text is a hard task; state-of-the-art methods in text classification achieve only modest performance in this domain[4]. In this area, some of the hardest problems involve acquiring large corpora tagged with detail linguistic expressions that indicate emotion.

In this study, we make an analysis on sentence emotion based on emotion words using Ren-CECps [a] (a Chinese emotion corpus developed by Ren-lab)[12]. Some classification methods (including C4.5 decision tree, SVM, NaiveBayes, ZEROR, and DecisionTable) have been compared. Then a supervised machine learning method (Polynomial kernel method) is proposed to recognize the eight basic emotions (Expect, Joy, Love, Surprise, Anxiety, Sorrow, Angry and Hate). Using Ren-CECps, we get the emotion lexicons for the eight basic emotions. Statistics show that the emotion lexicons derived from Ren-CECps are used more often in real use of language for emotional expressions than HOWNET sentimental lexicons. Polynomial kernel (PK) method is used to compute the similarities between sentences and the eight emotion lexicons. Then the experiential knowledge derived from Ren-CECps is used to recognize whether the eight emotion categories are present in a sentence. The experiments showed promising results compared with the above classification methods.

The remainder of this paper is organized as follows. Section 2 describes the construction of Ren-CECps. Section 3 presents the comparison of some classification methods (including C4.5 decision tree, SVM, NaiveBayes, ZEROR, and DecisionTable) for sentence emotion recognition based on emotion words. In section 4, Polynomial kernel method for sentence emotion recognition and experimental results are described. Section 5 concludes this study with closing remarks and future directions.

2. Introduction of Ren-CECps

With the increase in the number of people with access to the internet, and the availability of tools for easily creating a weblog, there has been a great increase in the number of weblogs. Weblogs are often touted as the new journalism. Writing suits the recording of facts and the communication of ideas, and their textual basis

[a] http://a1-www.is.tokushima-u.ac.jp/member/ren/Ren-CECps1.0/Ren-CECps1.0.html

makes them equally suitable for recording emotions and opinions. Blogs are selected as data source for Ren-CECps annotation.

Ren-CECps is constructed based on a relative fine-grained annotation scheme, annotating emotion in text at three levels: document, paragraph, and sentence. Sentence level is the basic level for emotion annotation, the annotation include intensities of the eight basic emotion classes, emotion holder/target, emotional keywords/phrases, rhetoric, emotional punctuation, emotional objective/subjective and emotional polarity.

Paragraph level is the upper level of sentence level, the annotation include intensities of the eight basic emotion classes, topic keywords to reflect the topic of a paragraph, and the numbers of topic sentences that can express the main points of this paragraph. Document level is the uppermost level in annotation; its annotation is similar to paragraph level. The tokenised text files are organized into XML documents, with Chinese segmentation tags and part-of-speech tags included as attributes of the tokens. An example document is listed in Fig. 1.

The main purpose of constructing this emotion corpus is to support the development and evaluation of emotion analysis systems in Chinese. The all dataset consisted of 1,487 blog articles published at sina blog, sciencenet blog, baidu blog, qzone blog, qq blog, and other blog websites. There are 11,255 paragraphs, 35,096 sentences, and 878,164 Chinese words contained in this corpus.

3. Sentence Emotion Analysis Based on Emotion Words

Sentences are the basic units for emotional expression. Previous work have included various approaches for sentence level emotion detection. According to the cues for emotion expression, there are two main methods for sentence emotion recognition: emotion provoking event based method and emotion words based method. Regarding the emotion words based method, which is seen as the most naive approach and probably also the most popular method.

The emotions of a sentence can be affected by many factors: emotion words, negative words, conjunctions, puntuations, contexts, and so on. To explore the role of emotion words for sentence emotion recognition, we compare five classic classification methods (including C4.5 decision tree, SVM, NaiveBayes, ZEROR, and DecisionTable) for sentence emotion recognition based on emotion words.

We select training corpus from Ren-CECps randomly. Training corpus includes 100 documents (containing 2,863 sentences). As sentence emotion is subjective entity, a sentence may evoke multiple emotions in different people's mind. A part of documents in Ren-CECps have been annotated by three annotators independently to measure agreement on the annotation of this corpus, which include 26 documents with a total of 805 sentences, 19,738 words. We select testing corpus from this part of corpus randomly. Testing corpus includes 10 documents (containing 387 sentences). An output of word emotion(s) will be regarded as a correct result if it is in agreement with any one item of word emotion(s) provided by the three anno-

```xml
<?xml version="1.0" encoding="GB2312" standalone="yes" ?>
- <document>
    <Joy>0.8</Joy>
    <Hate>0.0</Hate>
    <Love>0.9</Love>
    <Sorrow>0.4</Sorrow>
    <Anxiety>0.0</Anxiety>
    <Surprise>0.0</Surprise>
    <Anger>0.0</Anger>
    <Expect>0.0</Expect>
    <Topic>８０后 社会责任</Topic>
  - <title T="地震背后：解读８０后的社会责任">
      <Segmented_S>地震/v 背后/n ：/w 解读/v ８０/n 后/f 的/u 社会/n 责任/n</Segmented_S>
      <S_Length>15</S_Length>
      <Opinion_Fact>opinion</Opinion_Fact>
      <Polarity>中性</Polarity>
      <Joy>0.0</Joy>
      <Hate>0.0</Hate>
      <Love>0.0</Love>
      <Sorrow>0.0</Sorrow>
      <Anxiety>0.0</Anxiety>
      <Surprise>0.0</Surprise>
      <Anger>0.0</Anger>
      <Expect>0.0</Expect>
    </title>
  - <paragraph>
      <P_no>第1段情感标注</P_no>
      <Joy>0.0</Joy>
      <Hate>0.0</Hate>
      <Love>0.0</Love>
      <Sorrow>0.6</Sorrow>
      <Anxiety>0.0</Anxiety>
      <Surprise>0.0</Surprise>
      <Anger>0.0</Anger>
      <Expect>0.0</Expect>
      <Topic>沉思过往</Topic>
      <Summarize_Sentences>1</Summarize_Sentences>
    - <sentence S="静夜，沉思过往！">
        <S_no>第1段第1句标注</S_no>
        <Segmented_S>静夜/n ，/w 沉思/v 过往/v ！</Segmented_S>
        <S_Length>8</S_Length>
        <Keywords start="-1" position="3" end="-1" Surprise="0.0" Sorrow="0.3" POS="v" Opinionholder="0" Love="0.0" Joy="0.0" Hate="0.0" Expect="0.0" Anxiety="0.0" Anger="0.0">沉思
          </Keywords>
        <Opinion_Fact>opinion</Opinion_Fact>
        <Polarity>消极</Polarity>
        <Joy>0.0</Joy>
        <Hate>0.0</Hate>
        <Love>0.0</Love>
        <Sorrow>0.4</Sorrow>
        <Anxiety>0.0</Anxiety>
        <Surprise>0.0</Surprise>
        <Anger>0.0</Anger>
        <Expect>0.0</Expect>
      </sentence>
    - <sentence S="继汶川地震一篇建议书后再也没有更新博文，为逝去的同胞默哀、也为思考人生抑或生命的价值，包括我自己！">
        <S_no>第1段第2句标注</S_no>
        <Segmented_S>继/Vg 汶川/n 地震/v 一/m 篇/q 建议书/n 后/f 再也/d 没有/d 更新/v 博文/v ，/w 为/p 逝去/v 的/u 同胞/n 默哀/v 、/w 也/d 为/p 思考/v 人生/n 抑或/n 生命/n
          的/u 价值/n ，/w 包括/v 我自己/n ！</Segmented_S>
        <S_Length>49</S_Length>
        <Keywords start="-1" position="21" end="-1" Surprise="0.0" Sorrow="0.6" POS="v" Opinionholder="0" Love="0.0" Joy="0.0" Hate="0.0" Expect="0.0" Anxiety="0.0" Anger="0.0">逝去
          </Keywords>
        <Keywords start="-1" position="26" end="-1" Surprise="0.0" Sorrow="0.6" POS="v" Opinionholder="0" Love="0.0" Joy="0.0" Hate="0.0" Expect="0.0" Anxiety="0.0" Anger="0.0">默哀
          </Keywords>
        <Keywords start="-1" position="31" end="-1" Surprise="0.0" Sorrow="0.0" POS="v" Opinionholder="0" Love="0.0" Joy="0.0" Hate="0.0" Expect="0.0" Anxiety="0.4" Anger="0.0">思考
          </Keywords>
        <Keywords start="-1" position="40" end="-1" Surprise="0.0" Sorrow="0.0" POS="n" Opinionholder="0" Love="0.3" Joy="0.0" Hate="0.0" Expect="0.0" Anxiety="0.0" Anger="0.0">价值
          </Keywords>
        <noword start="-1" position="13" modifier_word_position="15" modifier_word_length="2" end="-1">没有</noword>
        <Opinion_Fact>opinion</Opinion_Fact>
        <punctuation E_type="焦虑">！</punctuation>
        <Polarity>消极</Polarity>
        <Joy>0.0</Joy>
        <Hate>0.0</Hate>
        <Love>0.0</Love>
        <Sorrow>0.7</Sorrow>
        <Anxiety>0.3</Anxiety>
        <Surprise>0.0</Surprise>
        <Anger>0.0</Anger>
        <Expect>0.0</Expect>
      </sentence>
    </paragraph>
  + <paragraph>
  + <paragraph>
  + <paragraph>
  + <paragraph>
  + <paragraph>
  </document>
```


Fig. 1. An annotated document in XML format.

tators. The accuracies are measured by F-value. The results of F-value include: (a) recognize emotion and unemotion sentences; (b) recognize the eight basic emotions for emotion sentences (complete matching); (c) recognize the eight basic emotions for emotion sentences (single emotion matching).

Table 1 gives the results of F-value for different classification methods for sentence emotion recognition based on emotion words.

Table 1. F-value for different classification methods for sentence emotion recognition based on emotion words.

(a) recognize emotion or unemotion sentences
(b) recognize the eight basic emotions for emotion sentences (complete matching)
(c) recognize the eight basic emotions for emotion sentences (single emotion matching)

Classification method	F-value		
	(a)	(b)	(c)
C4.5 decision tree	86.0	53.9	71.0
SVM	**98.0**	**84.5**	**95.7**
NaiveBayes	94.7	41.0	67.9
ZEROR	94.7	23.1	40.7
DecisionTable	94.7	23.1	40.7

As shown in Table 1, SVM showed the best classification results compared with other classification methods. We can find that the above classic classification methods are not satisfying for sentence emotion recognition when only use emotion word as feature.

4. Polynomial Kernel Method for Sentence Emotion Recognition

Kernel methods (KMs) have been well used for solving machine learning problems. KMs make use of the information encoded in the inner-product among all pairs of data items, avoiding explicitly to compute the feature vector for a given input. In this study, we explore polynomial kernel (PK) method for the problem of emotion recognition.

Based on Ren-CECps, we first use Polynomial kernel (PK) method to compute the similarities between sentences and emotion lexicons. Then the experiential knowledge derived from Ren-CECps is used to recognize whether the eight emotion categories (Expect, Joy, Love, Surprise, Anxiety, Sorrow, Angry and Hate) are present in a sentence.

4.1. *Polynomial kernel method*

KMs approach the problem by mapping the data into a high dimensional feature space, and a variety of methods can be used to find relations in the data in that space[13]. In the vector-space model, documents are represented by a matrix D, whose columns are the documents and rows are the terms. The corresponding kernel is given by the inner product between the feature vectors[14] , see Eq. (1) and Eq. (2).

$$K = D'D \tag{1}$$

$$K(d_1, d_2) =< \phi(d_1), \phi(d_2) >= \sum_{j=1}^{N} tf(t_j, d_1)tf(t_j, d_2) \tag{2}$$

Document d is represented by a row vector, see Eq. (3) .

$$\phi(d) = (tf(t_1, d), ..., tf(t_j, d), ..., tf(t_N, d),) \in R^N \tag{3}$$

Where $tf(t_j, d)$ is the frequency of term i appeared in document j.

A linear transformation is $\phi(d) = \phi(d) \times S$, where S is any appropriately shaped matrix, can be set by Eq. (4).

$$S = RP \tag{4}$$

Where R is a term-weight matrix, and is diagonal, whose entire $R(i, i)$ are the weight of the term i, can be defined by the inverse document frequency $idf(t) = ln(\frac{l}{df(t)})$, l is the total number of documents in the corpus, $df(t)$ is the number of documents that contain the given term. P is a term-document matrix, whose entire $P(i, j)$ are the emotion weight of the term i in document j. (The method of getting matrix P is described in 4.2)

The new kernel K for this feature space is defined by Eq. (5) .

$$K = D'D = (DS)'DS = (DRP)'DRP \tag{5}$$

For a given kernel $K(d_1, d_2)$, the derived polynomial kernel is defined by Eq. (6).

$$K(d_1, d_2) = (K(d_1, d_2) + c)^d \tag{6}$$

Fig. 2 shows the architectural overview of getting polynomial kernel (PK).

The text preprocessing and feature extraction process is Chinese word segmentation and stop words filtration. By vector space model (based on tf.idf computing), we can get the Term-Doc matrix and Tf.idf matrix. And then, we can get the eight emotion lexicons from the train corpus. With the emotional word intensity values and Term-Doc Matrix, we can get the emotion weight matrix. By kernel computing (Eq. (6)), the polynomial kernel (PK) is obtained.

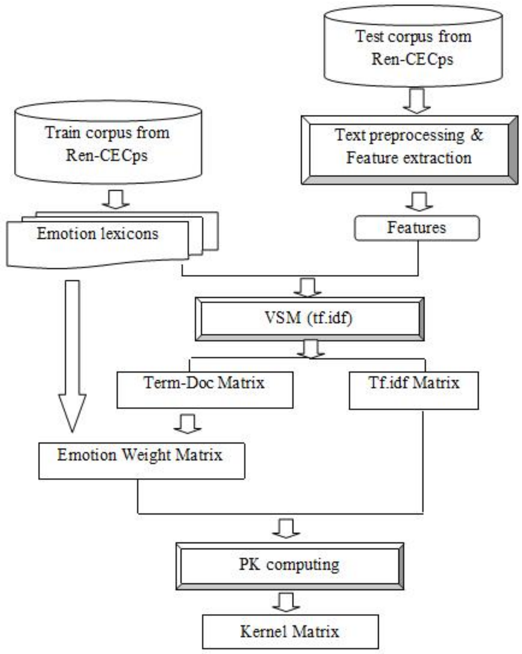

Fig. 2. The architectural overview of getting PK matrix.

4.2. *Emotional words annotation in Ren-CECps and emotion weight of terms*

In the annotation scheme of Ren-CECps, direct affective words and indirect affective words in a sentence are all annotated. Some emotional keywords may not bear emotions by themselves, but in a given context, they express emotions, and in similar contexts, they may express similar emotions.

An emotional keyword or phrase is represented as a vector to record its intensities of the eight basic emotional classes (Expect, Joy, Love, Surprise, Anxiety, Sorrow, Angry and Hate). For instance, the vector for the word "喜欢(like)" $\overrightarrow{w} = (0.0, 0.3, 0.9, 0.0, 0.0, 0.0, 0.0, 0.0)$ indicates the emotions of weak Joy and strong Love; "春天(English: spring)" $\overrightarrow{w} = (0.1, 0.3, 0.3, 0.0, 0.0, 0.0, 0.0, 0.0)$ indicates the emotions of weak Expect, Joy and Love. (The emotion classes and intensity values may be different because of different annotators). Emotion annotation is a hard task because the nature of emotion is inherently ambiguous (both in terms of the emotion classes and the natural language words that represent them). In the process of annotation, annotators were encouraged to follow their "first intuition".

Based on this corpus, we can get the emotion lexicons for the eight basic emotions. Table 2 is a comparison between the emotion lexicons derived from Ren-CECps (1487 articles) and the Chinese sentimental lexicons published by HOWNET (beta version)[15].

Table 2. Emotion lexicons comparison between Ren-CECps and HOWNET.

Statistics	Ren-CECps	HOWNET
Word num.	19,062	8,936
Word num. not appear in the opponent list	5,681	15,909
Total word num. that appear in the 1,487 articles	228,455	125,268
word num.(%) that the occurrence in the 1,487 articles below 2	13,006 (68.2%)	6,916 (77.4%)
word num.(%) that the occurrence in the 1487 articles from 2 to 5	2,556 ((13.4%))	722 (8.1%)
word num.(%) that the occurrence in the 1,487 articles above 5	3,500 (18.4%)	1704 (14.5%)

To the best of our knowledge, at present, there are no relatively large Chinese emotion lexicons. HOWNET sentimental lexicons are the only resources published for Chinese sentiment analysis. As shown in Table 2, the emotion lexicons derived from Ren-CECps are used more often in real use of language for emotional expressions than HOWNET sentimental lexicons. In HOWNET sentimental lexicons, 77.4% words of them only occur 0 or 1 time in this corpus; in Ren-CECps emotion lexicon, 31.8% words occur more than or equal 2 times. In the annotation scheme of Ren-CECps, the same word may be annotated with different emotions in different contexts, and the emotion intensity values may be different. For the eight basic emotions, their emotion lexicons include those words annotated with non-zero intensity value of the corresponding emotions. And the emotion intensity value for each word is the average value of the values in multiple annotations for the word. Then, we can get the matrix P (in section 4.1), whose entire $P(i,j)$ are the emotion weight of the term i in document j.

4.3. *Experiential Knowledge for Emotion Class Judgments*

Through statistical analysis, much experiential knowledge can be derived from Ren-CECps. We use the experiential knowledge of emotion number in each sentence to judge if a sentence contains some emotions. Table 3 shows the sentence numbers (percentage) containing emotion number from 0 to 8.

As can be seen from table 3, most sentences contain one or two emotions, few sentences contain more than four emotions. We also get the average value of emotion number for each sentence, which is 1.36. Based on this result, we can judge whether some emotions are present in a sentence.

Let $sim(s_i, l_j)$ is the similarity between $sentence_i$ and $lexicon_j$ computed by polynomial kernel (PK) method, and then we can get the similarity array from the

Table 3. Sentence numbers (percentage) containing emotion number from 0 to 8.

Emotion num.	0	1	2	3	4	5	6	7	8
Sentence num.	2,806	18,959	11,331	1,809	169	16	6	0	0
(%)	(8.0)	(54.0)	(32.3)	(5.2)	(0.5)	(0.0)	(0.0)	(0.0)	(0.0)

maximum down to the minimum, see Eq. (7).

$$\{sim(s_i, l_1), sim(s_i, l_2), ..., sim(s_i, l_8))\} \tag{7}$$

The emotions of $sentence_i$ are judged by Eq. (8).

$$Emotion(sentence_i) = \begin{cases} e(l_1), & if \quad sim(s_i, l_1) - sim(s_i, l_2) > \alpha \\ \{e(l_1), e(l_2)\}, & if \quad sim(s_i, l_1) - sim(s_i, l_2) \leq \alpha \\ & and \quad sim(s_i, l_2) - sim(s_i, l_2) > \beta \\ \{e(l_1), e(l_2), e(l_3)\}, & if \quad sim(s_i, l_1) - sim(s_i, l_2) \leq \alpha \\ & and \quad sim(s_i, l_2) - sim(s_i, l_3) \leq \beta \\ & and \quad sim(s_i, l_4) - sim(s_i, l_4) > \gamma \\ \{null\}, & if \quad sim(s_i, l_1) - sim(s_i, l_2) \leq \alpha \\ & and \quad sim(s_i, l_2) - sim(s_i, l_3) \leq \beta \\ & and \quad sim(s_i, l_4) - sim(s_i, l_4) \leq \gamma \end{cases} \tag{8}$$

in which, $e(l_i)$ is the emotion of $lexicon_i$, α, β, γ are threshold values to control the emotion number of a sentence.

4.4. *Experiments and results*

Ren-CECps (including 1487 articles) is used for the experiment, in which, 75% documents are used as training set, including 1115 documents (25,849 sentences) and the other 25% are used as test set, including 372 documents (9,247 sentences) . Table 4 shows the numbers of the eight emotion lexicons derived from Ren-CECps.

Table 4. Numbers of the eight emotion lexicons.

	Expect	Joy	Love	Surprise	Anxiety	Sorrow	Angry	Hate
Word num.	2,511	4,612	8,222	671	5,378	3,620	1,644	4,191

S1-S4 are examples of sentences and the emotions annotated by annotators.

S1): 我马上感觉到了她对女儿的思念之情。

English: I felt her strong yearnings toward her daughter right away.

Emotion (S1) = Love;

S2): 有多少人是快乐的呢？

English: How many people are happy?

Emotion (S2) = Anxiety, Sorrow;

S3): 她在同学中特别受欢迎。

English: She is greatly welcomed in her classmates.

Emotion (S3) = Love, Joy;

S4):这么美好的春光应该给人们带来温暖和欣慰，可是我的内心却冷冷作痛，这是为什么呢？

English: Such pleasant spring sunshine should bring people with warm and gratefulness, but I felt heartburn, why?

Emotion (S4) = Anxiety, Sorrow;

Table 5 shows examples of similarities between the eight emotion lexicons and sentences computed by polynomial kernel (PK) method. (The values of similarity are normalized.)

Table 5. Similarities between the eight emotion lexicons and sentences computed by PK method.

Sentences	S1	S2	S3	S4
Expect	0.13	0.10	0.11	0.11
Joy	0.16	0.13	0.19	0.16
Love	0.23	0.16	0.24	0.21
Surprise	0.05	0.05	0.05	0.05
Anxiety	0.16	0.18	0.13	0.14
Sorrow	0.12	0.13	0.09	0.11
Angry	0.06	0.09	0.07	0.08
Hate	0.09	0.13	0.11	0.12

In Eq. (8), when we set the threshold values $\alpha = 0.05$, $\beta = 0.03$, $\gamma = 0.01$, we get the emotions for sentence S1-S4:

Emotion (S1) = Love;

Emotion (S2) = Anxiety, Love, Sorrow;

Emotion (S3) = Love, Joy;

Emotion (S4) = Love, Joy, Anxiety;

From the above results, we can find that, the emotions computed by our method for S1 and S3 are in agreement with the emotion annotated by annotators. In S2, besides the two correct emotions of Anxiety and Sorrow, we get an error emotion of Love. In S4, we get two error emotions of Love and Joy.

Conducting a simple error analysis, we find that there is a word “快乐(happy)” in S2, which can express the emotion of Love; and there are more positive words in S4 such as “美好(pleasant)”, “温暖(warm)”, “欣慰(gratefulness)” cause the emotion of this sentence positive, , but more linguistic expressions are ignored, such as the conjunction “可是(but)” and rhetorical question.

F-value is used to evaluate this method. Our method obtained 65.0% F-value for recognize emotion or unemotion sentences, 46.7% F-value for recognize the eight basic emotions for emotion sentences (complete matching), and 70.1% F-value for recognize the eight basic emotions for emotion sentences (single emotion matching). When we use the same experimental setup as that have been described in Section 3, our method obtained 98.3% F-value for recognize emotion or unemotion sentences, 86.0% F-value for recognize the eight basic emotions for emotion sentences (complete matching), and 97.3% F-value for recognize the eight basic emotions for emotion sentences (single emotion matching), which are higher than SVM. The experiments showed promising results compared with the above classic classification methods (see section 3).

In conducting further error analysis, the main reasons include Chinese segmentation mistakes, ambiguous words, and conjunctions and collocations that express Expectation are not able to be detected. At the same time, contexts are important features should be considered.

5. Conclusions

The increased work has focused on emotion and sentiment analysis on written language, relevant researches including emotion recognition and classification, emotion prediction from text, sentiment classification and analysis, question answering based on sentimental analysis, summarization for multiple viewpoints. In this area, corpora are fundamental both for developing sound conceptual analysis and for training the emotion-oriented systems at different levels. Ren-CECps is a relative large emotion corpus annotated with detailed linguistic expressions for emotion in Chinese. It is developed to support the development and evaluation of emotion analysis systems.

Previous approaches to textual emotion analysis, have employed keyword spotting, lexical affinity, statistical methods, and handcrafted models. In this paper, we we make an analysis on sentence emotion based on emotion words using Ren-CECps . Some classification methods (including C4.5 decision tree, SVM, NaiveBayes, ZEROR, and DecisionTable) have been compared. Then a supervised machine learning method (Polynomial kernel method) is proposed to recognize the eight basic emotions (Expect, Joy, Love, Surprise, Anxiety, Sorrow, Angry and Hate). Using Ren-CECps, we get the emotion lexicons for the eight basic emotions. Statistics show that the emotion lexicons derived from Ren-CECps are used more often in real use of language for emotional expressions. Polynomial kernel (PK) method is used to compute the similarities between sentences and the eight emotion lexicons. Then the experiential knowledge derived from Ren-CECps is used to recognize whether the eight emotion categories are present in a sentence. The experiments showed promising results compared with the above classification methods.

Research into binary automatic sentiment classification can almost reach 90% accuracy under specific conditions[16]. However, classification of emotions seems more difficult than sentiment classification. The experiments also show that emotional words and phrases play important role for emotion recognition, but more linguistic expressions such as negative words, conjunctions, punctuations should be considered for more accurate recognition. In Ren-CECps, all of these are annotated, and then more work for emotion recognition will be done based on this corpus. We also find that experiential knowledge derived from this corpus is very useful, and more rules will be found.

6. Acknowledgments

We are grateful to Dr. Suzuki and Dr. Matsumoto for the helpful advice. This research has been partially supported by Ministry of Education, Science, Sprots and Culture, Grant-in-Aid for Challenging Exploratory Research, 21650030.

References

1. R. W. Picard. *Affective Computing*. The MIT Press, MA, USA, 1997.
2. F. Ren. *Recognizing Human Emotion and Creating Machine Emotion*, Invited Paper, Information, **8**(1), pp. 7-20, 2005 (Japanese).
3. F. Ren. *Affective Information Processing and Recognizing Human Emotion*, Electronic Notes in Theoretical Computer Science, **225**(1), pp. 39-50, 2009.
4. G. Mishne. *Experiments with Mood Classification in Blog Posts*, In Proceedings of the Style2005: The 1st Workshop on Stylistic Analysis of Text for Information Access, SIGIR 2005; Salvador, Brazil; Aug 15-19, 2005.
5. Y. Jung, H. Park and S. H. Myaeng, *A Hybrid Mood Classification Approach for Blog Text*, Lecture Notes in Computer Science, **4099**, pp.1099-1103, 2006.
6. H. Liu, H. Lieberman and T. Selker, *A model of textual affect sensing using real-world knowledge*, In Proceedings of the 2003 international conference on intelligent user interfaces, pp.125-132, 2003.
7. C. O. Alm, D. Roth and R. Sproat. *Emotions from Text: Machine Learning for Text-based Emotion Prediction*, In Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada, pp.579-586, 2005.
8. K. Lin, C. Yang and H.H. Chen. *What Emotions do News Articles Trigger in Their Readers?*, In Proceedings of the 30th Annual International ACM SIGIR Conference, poster, Amsterdam, Netherland, pp. 733-734, 2007.
9. A. Neviarouskaya, H. Prendinger and M. Ishizuka, *Textual Affect Sensing for Social and Expressive Online Communication*, In Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction. pp.218-229, 2007.
10. C. M. Lee and S. S. Narayanan. *Toward Detecting Emotions in Spoken Dialogs*, Journal of the American Society for Information Science. IEEE Trans. on Speech and Audio Processing **13**(2), pp.293-303, 2005.
11. S. Aman and S. Szpakowiczm, *Identifying Expressions of Emotion in Text*, Lecture Notes in Computer Science. **4629**, pp.196-205. 2007.
12. C. Quan and F. Ren, *Construction of a blog emotion corpus for Chinese emotional expression analysis*. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009), pp. 1446-1454, 2009.
13. T.D. Bie and N. Cristianini, *Kernel Methods for Exploratory Pattern Analysis: A Demonstration on Text Data. In Structural, Syntactic, and Statistical Pattern Recognition*, Lecture Notes in Computer Science, NJ: Springer Berlin / Heidelberg. **3138**, pp.16-29, 2004.

14. N. Cristianini, J. Shawe-Taylor and H. Lodhi. *Latent Semantic Kernels*, In Proceedings of the 18th International Conference on Machine Learning, Morgan Kaufman, pp.66-73, 2001.
15. Z. Dong and Q. Dong. *HowNet-a Hybrid Language and Knowledge Resource*, In Proceedings of Int'l Conf. Natural Language Processing and Knowledge Eng, pp.820-824, 2003.
16. T. Mullen and N. Collier. *Sentiment Analysis Using Support Vector Machines with Diverse Information Sources*, Proceedings of EMNLP, 2004.

Changqin Quan (Member)

She received the B.E, M.E. from Huazhong Normal University, Wuhan, China, in 2002 and 2005 respectively. She is currently a Ph.D. student in the faculty of engineering, the University of Tokushima, Japan. Her current research interests include Natural Language Processing and Affective Computing.

Fuji Ren (Member)

He received the Ph.D. Degree in 1991 from Faculty of Engineering, Hokkaido University, Japan. He worked at CSK, Japan, where he was a chief researcher of NLP from 1991. From 1994 to 2000, he was an associate professor in the Faculty of Information Sciences, Hiroshima City University. He became a professor in the faculty of engineering, the University of Tokushima in 2001. His research interests include Natural Language Processing, Artificial Intelligence, Language Understanding and Communication, and Affective Computing. He is a member of the CAAI, IEEJ, IPSJ, JSAI, AAMT and a senior member of IEEE, IEICE. He is a Fellow of The Japan Federation of Engineering Societies. He is the President of International Advanced Information Institute.