Building Emotion Lexicon from Weblog Corpora

Changhua Yang **Kevin Hsin-Yih Lin** **Hsin-Hsi Chen**
Department of Computer Science and Information Engineering
National Taiwan University
#1 Roosevelt Rd. Sec. 4, Taipei, Taiwan 106
{d91013, f93141, hhchen}@csie.ntu.edu.tw

Abstract

An emotion lexicon is an indispensable resource for emotion analysis. This paper aims to mine the relationships between words and emotions using weblog corpora. A collocation model is proposed to learn emotion lexicons from weblog articles. Emotion classification at sentence level is experimented by using the mined lexicons to demonstrate their usefulness.

1 Introduction

Weblog (blog) is one of the most widely used cybermedia in our internet lives that captures and shares moments of our day-to-day experiences, anytime and anywhere. Blogs are web sites that timestamp posts from an individual or a group of people, called bloggers. Bloggers may not follow formal writing styles to express emotional states. In some cases, they must post in pure text, so they add printable characters, such as ":-)" (happy) and ":-(" (sad), to express their feelings. In other cases, they type sentences with an internet messenger-style interface, where they can attach a special set of graphic icons, or emoticons. Different kinds of emoticons are introduced into text expressions to convey bloggers' emotions.

Since thousands of blog articles are created everyday, emotional expressions can be collected to form a large-scale corpus which guides us to build vocabularies that are more emotionally expressive. Our approach can create an emotion lexicon free of laborious efforts of the experts who must be familiar with both linguistic and psychological knowledge.

2 Related Works

Some previous works considered emoticons from weblogs as categories for text classification. Mishne (2005), and Yang and Chen (2006) used emoticons as tags to train SVM (Cortes and Vapnik, 1995) classifiers at document or sentence level. In their studies, emoticons were taken as moods or emotion tags, and textual keywords were taken as features. Wu et al. (2006) proposed a sentence-level emotion recognition method using dialogs as their corpus. "Happy, "Unhappy", or "Neutral" was assigned to each sentence as its emotion category. Yang et al. (2006) adopted Thayer's model (1989) to classify music emotions. Each music segment can be classified into four classes of moods. In sentiment analysis research, Read (2005) used emoticons in newsgroup articles to extract instances relevant for training polarity classifiers.

3 Training and Testing Blog Corpora

We select Yahoo! Kimo Blog[1] posts as our source of emotional expressions. Yahoo! Kimo Blog service has 40 emoticons which are shown in Table 1. When an editing article, a blogger can insert an emoticon by either choosing it or typing in the corresponding codes. However, not all articles contain emoticons. That is, users can decide whether to insert emoticons into articles/sentences or not. In this paper, we treat these icons as emotion categories and taggings on the corresponding text expressions.

The dataset we adopt consists of 5,422,420 blog articles published at Yahoo! Kimo Blog from January to July, 2006, spanning a period of 212 days. In total, 336,161 bloggers' articles were collected. Each blogger posts 16 articles on average.

We used the articles from January to June as the training set and the articles in July as the testing set. Table 2 shows the statistics of each set. On average, 14.10% of the articles contain emotion-tagged expressions. The average length of articles with tagged emotions, i.e., 272.58 characters, is shorter

[1] http://tw.blog.yahoo.com/

Proceedings of the ACL 2007 Demo and Poster Sessions, pages 133–136,
Prague, June 2007.

Table 1. Yahoo! Kimo Blog Emoticon Set.

ID	Emoticon	Code	Description	ID	Emoticon	Code	Description	ID	Emoticon	Code	Description	ID	Emoticon	Code	Description
1		:)	happy	11		:O	surprise	21		0:)	angel	31		(:|	yawn
2		:(	sad	12		X-(	angry	22		:-B	nerd	32		=P~	drooling
3		;)	winking	13		:>	smug	23		=;	talk to the hand	33		:-?	thinking
4		:D	big grin	14		B-)	cool	24		I-)	asleep	34		;))	hee hee
5		;;)	batting eyelashes	15		:-S	worried	25		8-)	rolling eyes	35		=D>	applause
6		:-/	confused	16		>:)	devil	26		:-&	sick	36		[-o<	praying
7		:x	love struck	17		:((	crying	27		:-$	don't tell anyone	37		:-<	sigh
8		:">	blushing	18		:))	laughing	28		[-(	not talking	38		>:P	phbbbbt
9		:p	tongue	19		:|	straight face	29		:o)	clown	39		@};-	rose
10		:*	kiss	20		/:)	raised eyebrow	30		@-)	hypnotized	40		:@)	pig

Table 2. Statistics of the Weblog Dataset.

Dataset	Article #	Tagged #	Percentage	Tagged Len.	Untagged L.
Training	4,187,737	575,009	13.86%	269.77 chrs.	468.14 chrs.
Testing	1,234,683	182,999	14.92%	281.42 chrs.	455.82 chrs.
Total	5,422,420	764,788	14.10%	272.58 chrs.	465.37 chrs.

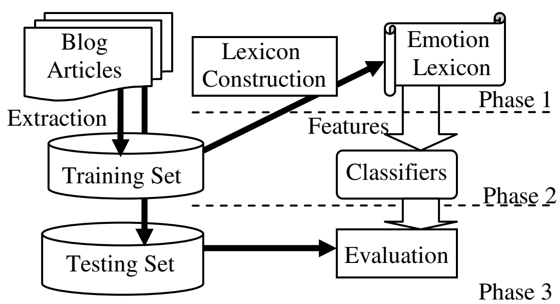

Figure 1. Emotion Lexicon Construction and Evaluation.

than that of articles without tagging, i.e., 465.37 characters. It seems that people tend to use emoticons to replace certain amount of text expressions to make their articles more succinct.

Figure 1 shows the three phases for the construction and evaluation of emotion lexicons. In phase 1, 1,185,131 sentences containing only one emoticon are extracted to form a training set to build emotion lexicons. In phase 2, sentence-level emotion classifiers are constructed using the mined lexicons. In phase 3, a testing set consisting of 307,751 sentences is used to evaluate the classifiers.

4 Emotion Lexicon Construction

The blog corpus contains a collection of bloggers' emotional expressions which can be analyzed to construct an emotion lexicon consisting of words that collocate with emoticons. We adopt a variation of pointwise mutual information (Manning and Schütze, 1999) to measure the collocation strength $co(e,w)$ between an emotion e and a word w:

$$co(e,w) = c(e,w) \times \log \frac{P(e,w)}{P(e)P(w)} \qquad (1)$$

where $P(e,w)=c(e,w)/N$, $P(e)=c(e)/N$, $P(w)=c(w)/N$, $c(e)$ and $c(w)$ are the total occurrences of emoticon e and word w in a tagged corpus, respectively, $c(e,w)$ is total co-occurrences of e and w, and N denotes the total word occurrences.

A word entry of a lexicon may contain several emotion senses. They are ordered by the collocation strength co. Figure 2 shows two Chinese example words, "哈哈" (ha1ha1) and "可惡" (ke3wu4). The former collocates with "laughing" and "big grin" emoticons with collocation strength 25154.50 and 2667.11, respectively. Similarly, the latter collocates with "angry" and "phbbbbt". When all collocations (i.e., word-emotion pairs) are listed in a descending order of co, we can choose top n collocations to build an emotion lexicon. In this paper, two lexicons (Lexicons A and B) are extracted by setting n to 25k and 50k. Lexicon A contains 4,776 entries with 25,000 sense pairs and Lexicon B contains 11,243 entries and 50,000 sense pairs.

5 Emotion Classification

Suppose a sentence S to be classified consists of n emotion words. The emotion of S is derived by a mapping from a set of n emotion words to m emotion categories as follows:

$$S \rightarrow \{ew_1,...,ew_n\} \xrightarrow[classification]{} \hat{e} \in \{e_1,...,e_m\}$$

哈哈 (ha1ha1) “*hah hah*”

Sense 1. (laughing) – ***co***: 25154.50

e.g., 哈哈… 我應該要出運了~

“*hah hah… I am getting lucky~*”

Sense 2. (big grin) – ***co***: 2667.11

e.g., 今天只背了單母音而已~哈哈

“*I only memorized vowels today~ haha*”

可惡 (ke3wu4) “*darn*”

Sense 1. (angry) – ***co***: 2797.82

e.g., 駭客在搞什麼…可惡

“*What's the hacker doing... darn it*”

Sense 2. (phbbbbt) – ***co***: 619.24

e.g., 可惡的外星人…

“*Damn those aliens*”

Figure 2. Some Example Words in a Lexicon.

For each emotion word ew_i, we may find several emotion senses with the corresponding collocation strength *co* by looking up the lexicon. Three alternatives are proposed as follows to label a sentence S with an emotion:

(a) Method 1

(1) Consider all senses of ew_i as votes. Label *S* with the emotion that receives the most votes.

(2) If more than two emotions get the same number of votes, then label *S* with the emotion that has the maximum *co*.

(b) Method 2

Collect emotion senses from all ew_i. Label S with the emotion that has the maximum *co*.

(c) Method 3

The same as Method 1 except that each ew_i votes only one sense that has the maximum *co*.

In past research, the approach used by Yang et al. (2006) was based on the Thayer's model (1989), which divided emotions into 4 categories. In sentiment analysis research, such as Read's study (2006), a polarity classifier separated instances into positive and negative classes. In our experiments, we not only adopt fine-grain classification, but also coarse-grain classification. We first select 40 emoticons as a category set, and also adopt the Thayer's model to divide the emoticons into 4 quadrants of the emotion space. As shown in Figure 3, the top-right side collects the emotions that are more positive and energetic and the bottom-left side is more negative and silent. A polarity classifier uses the right side as positive and the left side as negative.

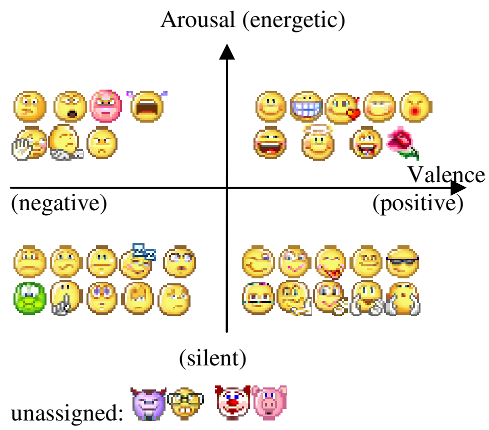

Figure 3. Emoticons on Thayer's model.

6 Evaluation

Table 3 shows the performance under various combinations of lexicons, emotion categories and classification methods. “Hit #” stands for the number of correctly-answered instances. The baseline represents the precision of predicting the majority category, such as “happy” or “positive”, as the answer. The baseline method's precision increases as the number of emotion classes decreases. The upper bound recall indicates the upper limit on the fraction of the 307,751 instances solvable by the corresponding method and thus reflects the limitation of the method. The closer a method's actual recall is to the upper bound recall, the better the method. For example, at most 40,855 instances (14.90%) can be answered using Method 1 in combination with Lexicon A. But the actual recall is 4.55% only, meaning that Method 1's recall is more than 10% behind its upper bound. Methods which have a larger set of candidate answers have higher upper bound recalls, because the probability that the correct answer is in their set of candidate answers is greater.

Experiment results show that all methods utilizing Lexicon A have performance figures lower than the baseline, so Lexicon A is not useful. In contrast, Lexicon B, which provides a larger collection of vocabularies and emotion senses, outperforms Lexicon A and the baseline. Although Method 3 has the smallest candidate answer set and thus has the smallest upper bound recall, it outperforms the other two methods in most cases. Method 2 achieves better precisions when using

Table 3. Evaluation Results.

	Baseline	Method 1 (M1)				Method 2 (M2)				Method 3 (M3)			
		Upp. R.	Hit #	Prec.	Reca.	Upp. R.	Hit #	Prec.	Reca.	Upp. R.	Hit #	Prec.	Reca.
Lexicon A 40 classes	**8.04%**	14.90%	14,009	4.86%	4.55%	14.90%	9,392	3.26%	3.05%	6.49%	13,929	4.83%	4.52%
Lexicon A Thayer	**38.38%**	48.70%	90,332	32.46%	29.35%	48.70%	64,689	23.25%	21.02%	35.94%	93,285	33.53%	30.31%
Lexicon A Polarity	**63.49%**	60.74%	150,946	54.25%	49.05%	60.74%	120,237	43.21%	39.07%	54.97%	153,292	55.09%	49.81%
Lexicon B 40 classes	8.04%	73.18%	45,075	15.65%	14.65%	73.18%	43,637	15.15%	14.18%	27.89%	45,604	**15.83%**	14.81%
Lexicon B Thayer	38.38%	89.11%	104,094	37.40%	33.82%	89.11%	118,392	**42.55%**	38.47%	63.74%	110,904	39.86%	36.04%
Lexicon B Polarity	63.49%	91.12%	192,653	69.24%	62.60%	91.12%	188,434	67.72%	61.23%	81.92%	195,190	**70.15%**	63.42%

Upp. R. – upper bound recall; **Prec.** – precision; **Reca.** – recall

Table 4. SVM Performance.

Method	Upp. R.	Hit #	Prec.	Reca.	F
Lexicon B M3	81.92%	195,190	70.15%	63.42%	66.62%
SVM 25 features	15.80%	38,651	79.49%	12.56%	21.69%
SVM 50 features	26.27%	62,999	77.93%	20.47%	32.42%
SVM 75 features	36.74%	84,638	74.86%	27.50%	40.23%
SVM 100 features	45.49%	101,934	72.81%	33.12%	45.53%
(Svm-25 + M3)	90.41%	196,147	70.05%	63.73%	66.74%
(Svm-50 + M3)	90.41%	195,835	70.37%	63.64%	66.83%
(Svm-75 + M3)	90.41%	195,229	70.16%	63.44%	66.63%
(Svm-100 + M3)	90.41%	195,054	70.01%	63.38%	66.53%

F = 2×(Precision×Recall)/(Precision+Recall)

Thayer's emotion categories. Method 1 treats the vote to every sense equally. Hence, it loses some differentiation abilities. Method 1 performs the best in the first case (Lexicon A, 40 classes).

We can also apply machine learning to the dataset to train a high-precision classification model. To experiment with this idea, we adopt LIBSVM (Fan et al., 2005) as the SVM kernel to deal with the binary polarity classification problem. The SVM classifier chooses top k (k = 25, 50, 75, and 100) emotion words as features. Since the SVM classifier uses a small feature set, there are testing instances which do not contain any features seen previously by the SVM classifier. To deal with this problem, we use the class prediction from Method 3 for any testing instances without any features that the SVM classifier can recognize. In Table 4, the SVM classifier employing 25 features has the highest precision. On the other hand, the SVM classifier employing 50 features has the highest F measure when used in conjunction with Method 3.

7 Conclusion and Future Work

Our methods for building an emotional lexicon utilize emoticons from blog articles collaboratively contributed by bloggers. Since thousands of blog articles are created everyday, we expect the set of emotional expressions to keep expanding. In the experiments, the method of employing each emotion word to vote only one emotion category achieves the best performance in both fine-grain and coarse-grain classification.

Acknowledgment

Research of this paper was partially supported by Excellent Research Projects of National Taiwan University, under the contract of 95R0062-AE00-02. We thank Yahoo! Taiwan Inc. for providing the dataset for researches.

References

Corinna Cortes and V. Vapnik. 1995. Support-Vector Network. *Machine Learning*, 20:273–297.

Rong-En Fan, Pai-Hsuen Chen and Chih-Jen Lin. 2005. Working Set Selection Using Second Order Information for Training Support Vector Machines. *Journal of Machine Learning Research*, 6:1889–1918.

Gilad Mishne. 2005. Experiments with Mood Classification in Blog Posts. *Proceedings of 1st Workshop on Stylistic Analysis of Text for Information Access.*

Jonathon Read. 2005. Using Emotions to Reduce Dependency in Machine Learning Techniques for Sentiment Classification. *Proceedings of the ACL Student Research Workshop*, 43-48.

Robert E. Thayer. 1989. *The Biopsychology of Mood and Arousal, Oxford University Press.*

Changhua Yang and Hsin-Hsi Chen. 2006. A Study of Emotion Classification Using Blog Articles. *Proceedings of Conference on Computational Linguistics and Speech Processing*, 253-269.

Yi-Hsuan Yang, Chia-Chu Liu, and Homer H. Chen. 2006. Music Emotion Classification: A Fuzzy Approach. *Proceedings of ACM Multimedia*, 81-84.

Chung-Hsien Wu, Ze-Jing Chuang, and Yu-Chung Lin. 2006. Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models. *ACM Transactions on Asian Language Information Processing*, 5(2):165-182.