2007 IEEE/WIC/ACM International Conference on Web Intelligence

Emotion Classification Using Web Blog Corpora

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

Abstract

In this paper, we investigate the emotion classification of web blog corpora using support vector machine (SVM) and conditional random field (CRF) machine learning techniques. The emotion classifiers are trained at the sentence level and applied to the document level. Our methods also determine an emotion category by taking the context of a sentence into account. Experiments show that CRF classifiers outperform SVM classifiers. When applying emotion classification to a blog at the document level, the emotion of the last sentence in a document plays an important role in determining the overall emotion.

1. Introduction

Emotion classification allows us to identify the feelings of individuals toward specific events. To use and evaluate the effectiveness of statistical learning methods for emotion classification, we need both a training dataset and a testing dataset. Weblog, or simply blog, is a very good dataset, which is collaboratively contributed by users, bloggers, on the web. As blogs become increasingly popular, the usage of emotion expressions is introduced. These expressions, which are commonly used in other Internet messaging interfaces, are now used by bloggers to share their experience with the blog community.

As more and more bloggers wish to share their feelings on blogs, some blog service providers start to allow their users to use non-verbal emotional expressions. Blog hosting sites such as LiveJournal[1] provide users with a special option to label their articles with their moods. Windows Live Spaces[2] and Yahoo! Kimo Blog [3] introduce an editor interface capable of letting users to insert system-defined emotion icons (called smileys or emoticons), which originate from instant messaging systems. People may insert emoticons into their text to show how they feel to make the blog content more vivid. Blog posts containing these emoticons form a readily available corpus for use in the classification of blog posts into emotion categories. Since thousands of new blog posts are created every day, the corpus formed by blog posts is expanding in size continuously without the need for costly professional annotators.

[1]http://www.livejournal.com/
[2]http://spaces.live/com/
[3]http://tw.myblog.yahoo.com/

2. Related Works

Some previous works used emoticons from blogs as categories for authors' sentiment classification. Mishne [6] used emoticons in LiveJournal posts to train a mood classifier at the document level. He used SVM [1] as the classifier and identified the intensity of the community mood [7]. Yang et al. [10] used Yahoo! Kimo Blog as corpora to build emotion lexicons. In their studies, emoticons were used to identify emotions associated with textual keywords. Lin et al. [4] focused on the classification of news articles into the readers' emotions instead of the authors'. Read [8] used text-based emoticons in newsgroup articles to extract instances relevant for training sentiment classifiers. Mao and Lebanon [5] trained CRF [3] classifiers on sequential sentiments with a movie review dataset.

The emotion information carried by a word can be treated as a kind of semantic orientation or subjectivity. Wiebe et al. [11] proposed a corpus labeled with subjectivities including opinions and emotions. Hatzivassiloglou and McKeown [2] asked human annotators to verify adjectives' semantic orientation. Turney and Littman [9] retrieved the web statistics of these terms. The obtained statistics were used to calculate other words' semantic orientations, which include both the polarity directions and the intensities.

3. System Framework

0-7695-3026-5/07 $25.00 © 2007 IEEE
DOI 10.1109/WI.2007.51

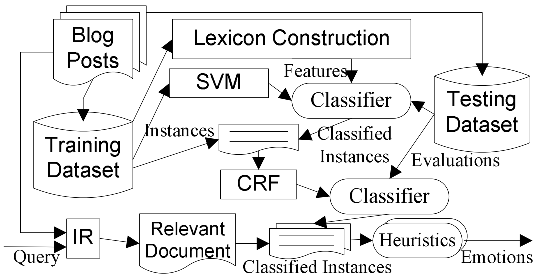

Figure 1. A Framework of Emotion Analysis.

Table 1. Statistics of the Blog Dataset.

Dataset (2006)	# of posts	# of tagged posts	Percentage	Avg. length of tagged	Avg. length of untagged
Training	4,176,250	575,009	13.77%	269.77 chrs.	468.14 chrs.
Testing	1,234,683	182,999	14.82%	281.42 chrs.	455.82 chrs.
Total	5,410,933	758,008	14.01%	272.58 chrs.	465.37 chrs.

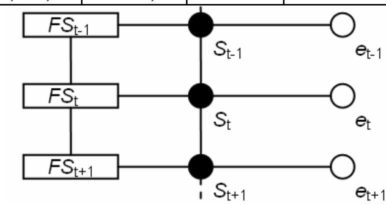

Figure 2. CRF Emotion Classifier.

Figure 1 illustrates our system's framework. Blog posts are collected and serve as the training and testing datasets. A lexicon construction method builds an emotion lexicon, which forms the fundamental features for emotion classification at the sentence level. SVM classifiers assign an emotion category to a sentence. Given a context (a sequence of sentences), CRF takes the context of a sentence into consideration when determining the appropriate emotion category. We then apply the trained classifiers to documents. We further propose heuristics to determine the emotion of the author.

We consider blog posts as datasets and the emoticons within as indicators for the emotions associated with sentences. The datasets consist of 5,410,933 blog posts published from Jan. to Jul., 2006 on Taiwan's Yahoo! Kimo Blog. Table 1 shows, on average, 14.01% of all posts contained tagged emoticons. We find that the average length of tagged posts, 272.58 characters, is shorter than that of untagged posts, 465.37 characters. It can imply that people use emoticons to replace certain portions of their text contents to make their articles more succinct.

4. Emotion Classification

Let a sentence S be composed of n terms $t_1, t_2, \ldots, t_n$. A feature set FS for S is defined as follows.

$$FS = \{t_k \mid t_k, 1 \le k \le n, t_k \in Lex\} \quad (1)$$

Those t_k ($1 \le k \le n$) appearing in the lexicon Lex mined by Yang et al. [10] are selected. Then an emotion

Table 2. Sample Emotion Set.

Positive Emotion			Negative Emotion		
Tagged Icon	Description	Sub Category	Tagged Icon	Descri-ption	Sub Category
	*laughing*	HAPPY		*crying*	SAD
	*big grin*			*sad*	
	*happy*				
	*rose*	JOY		*angry*	ANGRY
	*blushing*			*phbbbt*	
	*love struck*				

(E-1),(E-3) (for evaluating SVM and Bayesian),
$E1 \in$ {positive, negative}, $E3 \in$ {happy, sad, joy, angry}
S_1: $t_{11}t_{12}t_{13} \ldots t_{1|S1|} \rightarrow e_1 \in$ E1 or E3
S_2: $t_{21}t_{22}t_{23} \ldots t_{2|S2|} \rightarrow e_2 \in$ E1 or E3
…
S_n: $t_{n1}t_{n2}t_{n3} \ldots t_{n|Sn|} \rightarrow e_n \in$ E1 or E3

(E-2),(E-4) (for evaluating SVM and CRF)
$E2 \in$ {positive, negative}, $E4 \in$ {happy, sad, joy, angry}
S_1: $t_{11}t_{12} \ldots t_{1|S1|} \rightarrow e_1 \in E2$ or $E4$
Context$_1$
S_1: $e_1 \in$ {positive, negative}
S_2: (e_1) $t_{21}t_{22} \ldots t_{2|S2|} \rightarrow e_2 \in E2$ or $E4$
Context$_2$
S_2: $e_2 \in$ {positive, negative}
S_3: (e_2) $t_{31}t_{32} \ldots t_{3|S3|} \rightarrow e_3 \in E2$ or $E4$
…

Figure 3. Experiment Design.

assignment process is transformed to a classification problem:

$$S \underset{extraction}{\rightarrow} FS \underset{classification}{\rightarrow} \hat{e} \in \{e_1, \ldots, e_n\} \quad (2)$$

A set of features selected from a sentence is mapped to an emotion category. We adopt a version of SVM, LIBSVM[4], to train sentence emotion classifiers.

In addition to the features selected from a sentence, we incorporate the information of nearby sentences. Figure 2 sketches a context-based emotion classifier. To determine emotion e_t of sentence S_t, the new classifier considers the nearby context (features and emotion of S_{t-1}).

To train a sequential classification model for this problem, we use the MALLET[5] CRF implementation toolkit. A CRF training process aims to compute

$$p_\lambda(E \mid S) = \frac{1}{Z(S)} \exp\left(\sum_{t \in T} \sum_{k} \lambda_k f_k(E_{t-1}, E_t, S, t)\right) \quad (3)$$

while estimating the k weighing parameters λ_k for each feature f_k on transitions of time-sliced emotion status to maximize the conditional likelihood of $P_\lambda(E|S)$. Here E and S are random variables for emotions and sentences, respectively. Then the classification problem defined in (9) with context can be modified as:

[4] http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[5] http://mallet.cs.umass.edu/

$$S_1,...,S_n \underset{extraction}{\rightarrow} FS_1,...,FS_n \underset{classification}{\rightarrow} \hat{e}_1,...\hat{e}_n \quad (4)$$

Given n sentences at one time, the corresponding emotions will be returned at the same time. The sentence emotion classifiers' results provide the CRF method the context information necessary to enhance the prediction accuracy.

We design three criteria to classify a document into an emotion category:

(c1) assign the emotion that appears in the most number of the classified sentences,

(c2) assign the emotion that appears in the longest series of classified sentences that have the identical emotions, and

(c3) assign the emotion of the last classified sentence.

For example, a document may be classified as *happy* because it consists of happy sentences primarily; has continuously emphasized happy sentences; or the author expressed that he was happy by his final statement with a happy emoticon.

5. Experiment Setup

Table 2 shows the emotion taxonomy used in our experiments. There are two major emotion categories – positive and negative. The positive emotion category is further partitioned into two subcategories HAPPY and JOY. Similarly, the negative emotion category has subcategories SAD and ANGRY. Figure 3 shows four sets of experiments. Coarser-grained emotion categories, i.e., positive and negative, are adopted in experiments (E-1) and (E-2), while finer-grained emotion categories of HAPPY, JOY, SAD, and ANGRY, are adopted in experiments (E-3) and (E-4).

In addition to the SVM and CRF methods, a Bayesian classifier is used for performance comparison with the SVM. This classifier employs a conditional probability of each word in the mined emotion lexicon. The document-level emotion classifier is defined as follows. After processing 2,586 testing documents using sentence-level SVM classifiers and context-level CRF classifiers, each document contains emotion tagging on possible sentences. The three criteria, (c1), (c2) and (c3), are then applied and an emotion category for each document is proposed.

6. Experimental Results

We use different numbers of emotion keywords as features to train emotion classifiers. The classifiers are applied to the testing datasets. We report the precision, recall and F-Score. The total number of testing instances of experiments E-1 and E-3 is 31,255 and that of experiments E-2 and E-4 is 17,887.

Table 3. (E-1) on Coarser Emotion Categories.

Experiment Setup	Precision	Recall	F-Score
Bayesian 50 features	78.30%	31.38%	44.80%
Bayesian 100 features	74.66%	42.11%	53.85%
Bayesian 150 features	73.78%	46.72%	57.21%
SVM 50 features	78.67%	31.21%	44.69%
SVM 100 features	75.02%	41.67%	53.58%
SVM 150 features	74.02%	46.78%	57.33%

Table 4. (E-2) on Coarser Emotion Categories.

Experiment Setup	Precision	Recall	F-Score
SVM 50 features	78.22%	15.56%	25.96%
SVM 100 features	74.18%	26.33%	38.86%
SVM 150 features	72.41%	32.24%	44.61%
CRF 50 features	82.27%	16.37%	27.31%
CRF 100 features	80.34%	28.51%	42.09%
CRF 150 features	79.87%	35.56%	49.21%

Table 5. (E-3) on Finer Categories.

Experiment Setup	Precision	Recall	F-Score
Bayesian 50 features	51.30%	20.34%	29.13%
Bayesian 100 features	48.89%	27.13%	34.89%
Bayesian 150 features	48.41%	30.56%	37.47%
Bayesian 500 features	47.54%	37.73%	42.07%
SVM 50 features	53.92%	21.40%	30.64%
SVM 100 features	49.94%	27.74%	35.67%
SVM 150 features	48.04%	30.36%	37.21%
SVM 500 features	44.67%	35.49%	39.56%

Table 6. (E-4) on Finer Emotion Categories.

Experiment Setup	Precision	Recall	F-Score
SVM 50 features	48.83%	9.72%	16.21%
SVM 100 features	45.09%	16.00%	23.62%
SVM 500 features	40.94%	27.47%	32.88%
CRF 50 features	56.00%	11.14%	18.59%
CRF 100 features	54.74%	19.43%	28.68%
CRF 500 features	53.98%	36.22%	43.35%

Table 3 summarizes the performance of sentence-based SVM and Bayesian classifiers when using 50, 100, and 150 emotion words as features. SVM classifiers outperform the Bayesian classifiers in almost all cases. As we increase the number of keywords used as features, the precision drops gradually. However, recall and F-Score increase. Table 4 shows that the CRF classifier performs better when the emotion category of the previous sentence is used.

When four subcategories of emotion – HAPPY, JOY, ANGRY and SAD – are used, we increase the maximum feature set size to 500. As shown in Table 5, the introduction of more feature words benefits both the Bayesian and SVM classification's performances. However, SVM seems to suffer from an overfitting problem. With more than 150 keywords incorporated as features, the Bayesian classifiers perform better than SVM classifiers. Table 6 further shows when the previous sentences' emotions are known, CRF can achieve better than SVM while still maintaining precisions over 53%.

For document-level emotion classification tasks, 50.39% of the documents are positive. It can be treated as the baseline precision for the document-level emotion classification experiment. Table 7 shows the

Table 7. Document-Level Evaluation.

Heuristic	Precision	Recall	F-Score
(c1)	57.55%	54.80%	56.14%
(c2)	58.03%	55.45%	56.71%
(c3)	59.23%	58.58%	58.90%

Table 8. Emotion Dist. of Nearby Sentences.

Previous Current	Pos. Pos.	Pos. Neg.	Neg. Pos.	Neg. Neg.	Total Inst.
Training Distribution	28,347 (55.8%)	5,492 (10.8%)	6,008 (11.8%)	10,960 (21.6%)	50,807
Testing Distribution	9,885 (55.3%)	2,035 (11.4%)	2,213 (12.4%)	3,754 (22.0%)	17,887
SVM-150 Ans. Dist.	5,261 (66.1%)	120 (1.5%)	1,932 (24.3%)	650 (8.2%)	7,963+
CRF-150 Ans. Dist.	5,060 (63.5%)	321 (4.0%)	885 (11.1%)	1,697 (21.3%)	7,963+

+ Among 17,887 testing instances, only 7,963 instances containing at least one of 150 feature keywords are classifiable in these models.

performance of the three proposed criteria. All reported precisions are higher than the baseline. The heuristic (c3) performs the best and implies that bloggers tend to emphasize their feelings in the last sentence such as attach a *happy* emoticon with a concluding sentence "Finally, I pass the exam."

7. Discussion

This section discusses the effects of different classifiers and the collaborative annotation of training data set. The experiments show that CRF classifiers outperform SVM classifiers. This means that the emotional information of nearby sentences affect each other so that a context-aware classifier is useful. CRF has the ability to learn the transition of emotion categories from one sentence to the next, but SVM cannot do that very well. The category of the previous sentence can be used as a feature in SVM, but it does not guarantee that SVM can learn the transition relationship.

Table 8 illustrates the emotion relationships in contexts further. The second and third rows of Table 10 list the emotion distribution of two continuous sentences in the experiment (E-2). The distribution of the training dataset is similar to that of the testing dataset. Besides, sentences of the same polarity follow each other dominate the cases. Positive→Positive and Negative→Negative occupy 55.8% (55.3%) and 21.6% (22.0%) respectively, in the training (testing) set. It shows bloggers express consistent emotions within a context, and most bloggers express positive emotions. In this way, SVM classifiers tend to put more instances into the positive categories. The 4th row of Table 10 verifies the point. The SVM classifier predicts almost 90% of the instances as positive. In contrast, CRF can deal with more negative→negative cases. The 5th row of Table 10 shows the distribution of outcome of a CRF classifier is more similar to that of the source datasets than that of the SVM classifier.

8. Conclusion and Future Work

In this paper, we perform emotion classification on web blog corpora. The experiments show that the CRF classifier performs better than the SVM classifier at the sentence level. At the document level, the strategy of picking the last sentence's emotion as the answer outperforms all other strategies. It implies that humans summarize their emotion of the overall document by the last sentence of the document.

Blog corpora make supervised emotion classification feasible. In this paper, we use words as features. Feature coverage is an important issue. Instances without features are unclassifiable at the current stage. Other types of features, ways to find feature association and to increase the recall of classification need further investigation.

Acknowledgements

Research of this paper was partially supported by Excellent Research Projects of National Taiwan University under the contract 95R0062-AE00-02.

References

[1] C. Cortes and V. Vapnik, "Support-Vector Network", *Machine Learning*, 20, pp. 273–297, 1995.

[2] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives", *Proceedings of the ACL-EAACL*, pp. 174-181, 1997.

[3] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", *Proceedings of International Conference on Machine Learning*, pp. 282-289, 2001.

[4] K. H.-Y. Lin, C. Yang, and H.-H. Chen, "What Emotions News Articles Trigger in Their Readers?", *Proceedings of SIGIR*, pp. 733-734, 2007.

[5] Y. Mao and G. Lebanon, "Sequential Models for Sentiment Prediction", *Proceedings of ICML workshop on Learning in Structured Output Spaces, 2006*.

[6] G. Mishne, "Experiments with Mood Classification in Blog Posts", *Proceedings of 1st Workshop on Stylistic Analysis of Text for Information Access*, 2005.

[7] G. Mishne and M. de Rijke, "Capturing Global Mood Levels using Blog Posts", *Proceedings of AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*, pp. 145-152, 2006.

[8] J. Read, "Using Emotions to Reduce Dependency in Machine Learning Techniques for Sentiment Classification", *Proceedings of the ACL Student Research Workshop*, pp. 43-48, 2005.

[9] P. D. Turney and M. L. Littman, "Measuring Praise and Criticism: Inference of Semantic Orientation from Association", *ACM Transactions on Information Systems*, 21, 4, pp. 315-346, 2003.

[10] C. Yang, K. H.-Y. Lin, and H.-H. Chen, "Building Emotion Lexicon from Weblog Corpora", *Proceedings of the 45th Annual Meeting of ACL*, pp. 133-136, 2007.

[11] J. Wiebe, T. Wilson, and C. Cardie, "Annotating Expressions of Opinions and Emotions in Language", *Language Resources and Evaluation*, 39, 2, 3, pp. 165-210, 2005.