Determining Mood for a Blog by Combining Multiple Sources of Evidence

Yuchul Jung, Yoonjung Choi and Sung-Hyon Myaeng
School of Engineering, Information and Communications University
119, Munjiro, Yuseong-gu, Daejeon, 305-732, South Korea
{enthusia77, choiyj35, myaeng}@icu.ac.kr

Abstract

Mood classification for blogs is useful in helping user-to-agent interaction for a variety of applications involving the web, such as user modeling, recommendation systems, and user interface fields. It is challenging at the same time because of the diversity of the characteristics of bloggers, their experiences, and the way moods are expressed. As an attempt to handle the diversity, we combine multiple sources of evidence for a mood type. Support Vector Machine based Mood Classifier (SVMMC) is integrated with Mood Flow Analyzer (MFA) that incorporates commonsense knowledge obtained from the general public (i.e. ConceptNet), the Affective Norms English Words (ANEW) list, and mood transitions. In combining the two different approaches, we employ a statistically weighted voting scheme based on the Support Vector Machine (SVM). For evaluation, we have built a mood corpus consisting of manually annotated blogs, which amounts to over 4000 blogs. Our proposed method outperforms SVMMC by 5.68% in precision. The improvement is attributed to the strategy of choosing more trustable classification results in an interleaving fashion between the SVMMC and our MFA.

1. Introduction

A blog is a web site, where anybody can write about his or her own personal experiences and thoughts on a voluntary basis. As such, it is likely to reflect the user's personality and cultural biases, sometimes forming a unique society. Although mood classification in blogs is considered meaningful in helping user-to-agent interactions, like user modeling, recommendation systems, and user interface fields, compared to topicality-based classification of text, it is quite challenging and difficult in many aspects.

A recent approach [1] to mood classification of text used Support Vector Machine (SVM) with 6 features: frequency counts, length, semantic orientation, Point-wise Mutual Information for Information Retrieval (PMI-IR), emphasized words, and special symbols. The small improvement (average 8%) over a low baseline (50%) seems to be a limit due to the inability to deal with idiosyncratic natures of moods and blogs. For example, although an author is under a certain mood when starting to write a blog document, the initial mood may not be maintained all the way to the end. Some blogs are so intertwined that even human readers would have difficulty in identifying the mood, not to mention the statistically motivated method using surface level features. That is, we can hardly say that the SVM classifier is potent in the mood classification of blogs all the time.

Currently, there exist some commonsense knowledge and affective linguistic resources that could contribute to mood classification, provided that they are combined selectively.

- For instance, ConceptNet [2], a huge commonsense knowledge base, supports many practical textual-reasoning tasks over real-world documents including topic-jisting, affect-sensing, analogy-making, and other context-oriented inferences. The knowledge base is a semantic network presently consisting of over 1.6 million assertions of commonsense knowledge covering the spatial, physical, social, temporal, and psychological aspects of everyday life.

- The ANEW list [3], created from a psychological study, contains 1,034 unique terms with affective pleasure (unpleasant ~ pleasant), arousal (calm ~ excited), and dominance (submissive ~ dominance) scores. It can be used to identify different mood types based on lexical analysis by mapping terms in text to those in the list.

To detect the mood of blog text more accurately, a hybrid approach to mood classification was developed [4]. The approach incorporates the SVM classifier, the GuessMood function of ConceptNet [2], and the

0-7695-3026-5/07 $25.00 © 2007 IEEE
DOI 10.1109/WI.2007.140

Affective Norms English Words (ANEW) list [3]. However, the results show that this approach is not is not significantly better than SVM, and that its mood corpus, data being used to evaluate the approaches, is not a gold standard.

Motivated by the previous work, we have employed a method for combining multiple sources of evidence to detect the mood associated with blogs. As a meaningful basic source, we use the SVM classifier. In addition, to observe fluctuations of moods between several paragraphs in the given blog text, a mood flow analyzer was developed by utilizing a revised version of GuessMood and an ANEW based affective sensing module.

Since we found that the two different approaches complement each other in the mood classification performance aspects, we have designed and tested a class-based weighted voting, a heuristic weighted voting, and a statistically weighted voting to combine the two classifiers for 4 mood types (happy, sad, angry, and fear). For more robust evaluation, we have built a manually annotated mood corpus which amounts to over 4000 blogs. Among the three weighted voting schemes, the statistical weighted voting method showed the best performance with an average of 93.31% in precision, which outperforms the SVMMC by 5.68% when the gold standard was used for testing.

2. A Blog Mood Corpus

"LiveJournal.com" is open to anyone like other blog services, making it easy to share opinions & thoughts and to join. A very useful option in the site is that a user can choose one mood among the predefined 132 moods, such as cheerful, tired, sad, angry, etc.

We have collected over 50GB of blog texts from "LiveJournal.com" semi-automatically by mapping our mood categories (happy, sad, angry, and fear) to appropriate "Mood IDs" referring to the LiveJournal's mood hierarchy[1].

Prior to using human labor in the corpus annotation process, 14000+ blog documents were extracted from the 50GB blog texts using a keyword spotting technique to select promising mood containing blogs. In addition, to avoid irregular, meaningless, and ambiguously long blog documents, it ignores blog texts whose length is less than 5 or more than 40 sentences.

As a result, we have 7,189 blog documents as a gold standard for mood classification. The gold standard was built manually by two people who annotated individual documents and ensured inter-coder agreements. To collect a sufficient numbers of documents for each mood type, we have run our crawler 20 times and conducted subsequent mood annotations for four months. As a result, we have acquired 1000+ documents for each mood.

3. Proposed System

Our system consists of two different mood classification mechanisms: the SVM based Mood Classifier (SVMMC) and the Mood Flow Analyzer (MFA). When a blog document comes in, the system goes through a specially designed preprocessing step for blog texts to generate features for the two different classification methods. It includes handling broken words, eliminating stop-words, and normalizing onomatopoeic words and blog-specific characteristics (emoticons, abbreviated words, etc) (Fig. 1). The two different classifiers carry out their own strategies before making a final decision. In the final resolver, we employ a statistically weighted voting scheme to select a mood considering contents of the inputted blog text.

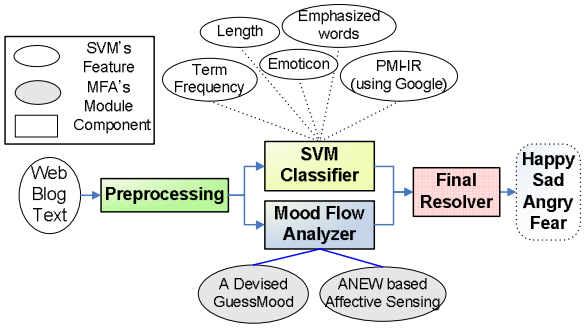

Figure 1. Overall architecture

3.1. SVM based Mood Classifier (SVMMC)

Using the 5 features in Fig. 1, i.e., length, term frequency, emotion, emphasized words, and PMI-IR, the system performs SVM[2] based mood classification to assign a mood category according to the inputted blog text. It is a straightforward application of the classifier to the obtained data.

3.2. Mood Flow Analyzer (MFA)

As in Fig. 2, our MFA segments a blog document into paragraphs. After that, in the paragraph analysis, the number of mood terms is counted to select a scheme between the revised GuessMood and the ANEW based affective sensing module. If the number of mood terms is bigger than the experimental threshold (i.e. 3), the latter is chosen. It can be

[1] http://www.livejournal.com/moodlist.bml?moodtheme=140&mode=tree

[2] http://svmlight.joachims.org/

effectively applied to usual writer's blog text where affective words appear explicitly. For the GuessMood function, ConceptNet data were specially reorganized by extracting only affection related sentences from the OMCS raw corpus for time-efficiency and performance. Our ANEW based affective text sensing module has been built by leveraging a model for affective text sensing approach [2]. If a mood is sustained without transitions throughout the whole blog document, the "Mood Resolver" module only checks the consistency and assigns the mood as the final result. When some paragraphs have different mood types, heuristically measured weights are added into the results of paragraph analysis regarding the position of a paragraph in the document. Only when the paragraph should be emphasized (especially, starting/ending part), some heuristic weights obtained through several thousands of trials with our blog corpus are applied in the mood resolver function.

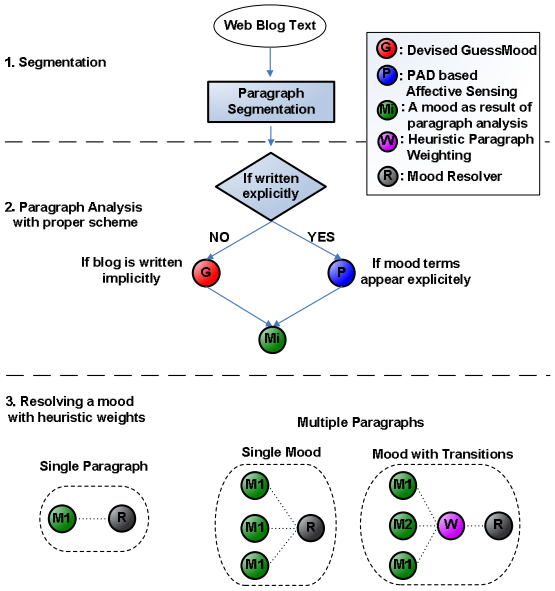

Figure 2. Details of mood flow analyzer (MFA)

3.3. Final Resolver

Table 1. Blogs statistics: the MFA succeeded but the SVMMC failed

Category	Happy	Sad	Angry	Fear	Total
Total number of blog texts	1125	1031	1136	1011	4303
Only the MFA is Correct	37	78	40	91	246
Percentage	3.29%	7.57%	3.52%	9.00%	5.72%

Because we observed that our MFA had correctly classified moods of about 246 blogs (5.72% of the blog corpus) that the SVMMC failed (Table 1), we use its classification results when its statistically obtained confidence level is higher than the SVMMC's results

To calculate the confidence level initially, we have tested a class-based weighting scheme [5], but its performance was almost the same as that of the SVMMC because the SVMMC is superior to the MFA in every mood classification. As an alternative, we have tried a heuristically weighted voting which concerns class level preference on classifiers and some statistical information such as frequencies of affective keywords appeared in blog texts and leading affective keywords' tendency toward a specific mood. However, this scheme has merely shown small progresses on each mood classification. Through those two trials failed in identifying when the system should believe the MFA's results or the SVMMC's results, we recognized that achieving high performance in weighted voting is almost impossible without consideration of the contents of documents the SVMMC failed but MFA succeeded.

In order to make a statistically significant weighted voting scheme that is aware of contents of blog texts, we have implemented a statistical classifier trained with the SVMMC-preference clues (blog texts both the MFA and the SVMMC correctly classified and showed same decisions) and the MFA-preference clues (blog texts the MFA successfully classified but the SVMMC failed). To accommodate diverse features of the two training data and maximize preference discrimination based on the contents of blog texts, SVM machine learning is used and a TF*IDF weighting [6] is adopted for feature weighting.

As in Fig. 4, the final resolver depends on the weights obtained by the statistical classifier's decision on the inputted blog text. The final resolver produces a mood chosen between the MFA and the SVMMC considering the weights.

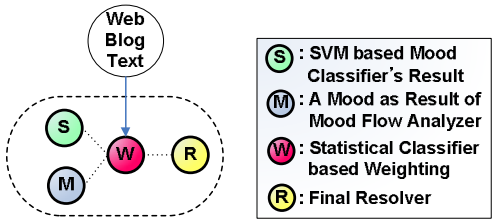

Figure 4. Details of the final resolver

4. Experiments and Discussion

The main goal of the experiments is to evaluate how much improvement the combining approach can make over each of the different methods (Table 2). A total of manually chosen 4,303 blog documents (happy: 1,125, sad: 1,031, angry: 1,036, and fear: 1,011) from our mood corpus mentioned in Section 2 were used for training and testing, respectively.

Table 2. Mood classification statistics

Category (Total #)	M.-1 (Prec.)	M.-2/C.-1 (Prec.)	C.-2 (Prec.)	C.-3 (Prec.)
Happy (1125)	975 (86.61%)	1080 (95.97%)	1082 (96.18%)	1116 (99.22%)
Sad (1031)	470 (45.59%)	816 (79.15%)	830 (80.50%)	894 (86.71%)
Angry (1136)	721 (63.47%)	1049 (92.34%)	1067 (93.93%)	1089 (95.86%)
Fear (1011)	651 (64.39%)	826 (81.70%)	836 (82.69%)	916 (90.60%)
Total (4303)	2817 (65.46%)	3771 (87.63%)	3815 (88.66%)	4015 (93.31%)

Method-1 (Mood flow analyzer): Manually selected training data helps in constructing a classifier with limited coverage, so its performance is 65.46% on average.

Method-2 (SVM based classifier) / Combination-1 (Class-based weighted voting): When the SVM classifier was tested with the selected testing data, 5-fold cross validation was taken to get average precision of the SVM classifier in mood classification. Its high precision (average of 87.63%) shows that our corpus is highly refined and can be used as a trustable gold standard. Because SVMMC outperforms MFA in every mood, the result of the class-based weighted voting based combination, which chooses the best performing classifier at class level, is almost the same as SVMMC.

Combination-2 (Combination using a heuristic voting): This experiment is intended to catch potentially existing preferences on one of two classifiers (the MFA and the SVMMC) by just using each classifier's mood classification results and their correct moods annotated by human. In every testing with this combination approach, classification accuracy is slightly over that of the SVMMC. In addition, the method is very ad hoc and needs time consuming observation on the overall corpus.

Combination-3 (Combination using a statistically weighted voting): To overcome C.-2 approach's flaws, we have built the "Final Resolver" in which a statistical classifier is embedded to give weights on the most trustable classifier between the MFA and the SVMMC. With the help of the final resolver's high discrimination capability, we obtained very meaningful enhancements over the 4 moods (*happy*: 3.24%, *sad*: 7.57%, *sad*: 3.52%, *fear*: 8.90%, and average 5.68%) when compared with the Method-2.

5. Conclusion

This paper presents a statistically weighted voting scheme that is applicable to the two different classifiers – the Mood Flow Analyzer (MFA) and the SVM based Mood Classifier (SVMMC) - for mood classification of blog texts. To facilitate each classifier's advantages, the scheme's weights are calculated and applied based on the content of blogs. In addition, a manually annotated mood corpus (i.e. blogs) has been built and used to evaluate our proposed model. Its accuracy was as high as 93.31% in precision on average, which is about 5.68% higher than the SVMMC.

Mood classification of blog texts is a difficult task without human-level intelligence because of diverse situations the author were in and expressions they use for their mood. However, we identified that the content-aware, statistically weighted voting scheme used in combining two different classifiers works well for mood classification of blog texts.

6. Acknowledgements

This work is supported by the Korea Foundation for International Cooperation of Science & Technology (KICOS) through a grant provided by the Korean Ministry of Science & Technology (MOST) in K20711000007-07A0100-00710.

7. References

[1] Mishne, G. Experiments with Mood Classification in Blog Posts. *Style2005 at SIGIR 2005*, 2005.

[2] Liu, H. and Singh, P. ConceptNet – a practical commonsense reasoning tool-kit, *BT Technology Journal*: 211-226, 2004.

[3] Bradley, M.M. and Lang, P. J. Affective norms for English words (ANEW). Gainesville, FL. The NIMH Center for the Study of Emotion and Attention, University of Florida, 1999.

[4] Jung, Y.C., Park, H.G., and Myaeng, S.-H. A Hybrid Mood Classification Approach for Blog Text, *PRICAI2006*, pp. 1099-1103.

[5] Yanmin, S., Mohamed S. K., and Andrew, K.C. Wong. Empirical Study on Weighted Voting Multiple Classifiers, *ICAPR 2005*, LNCS 3686, pp. 335-344, 2005.

[6] Baeza-Yates R., and Ribeiro B., *Modern Information Retrieval*, Addison Wesley, 1998.