Linguistic Template Extraction for Recognizing Reader-Emotion and Emotional Resonance Writing Assistance

Yung-Chun Chang[1,2], **Cen-Chieh Chen**[1,3], **Yu-Lun Hsieh**[1,3], **Chien Chin Chen**[2], **Wen-Lian Hsu**[1*]
[1]Institute of Information Science, Academia Sinica, Taipei, Taiwan
[2]Department of Information Management, National Taiwan University, Taipei, Taiwan
[3]Department of Computer Science, National Chengchi University, Taipei, Taiwan
[1]{changyc,can,morphe,hsu}@iis.sinica.edu.tw, [2]patonchen@ntu.edu.tw

Abstract

In this paper, we propose a flexible principle-based approach (PBA) for reader-emotion classification and writing assistance. PBA is a highly automated process that learns emotion templates from raw texts to characterize an emotion and is comprehensible for humans. These templates are adopted to predict reader-emotion, and may further assist in emotional resonance writing. Results demonstrate that PBA can effectively detect reader-emotions by exploiting the syntactic structures and semantic associations in the context, thus outperforming well-known statistical text classification methods and the state-of-the-art reader-emotion classification method. Moreover, writers are able to create more emotional resonance in articles under the assistance of the generated emotion templates. These templates have been proven to be highly interpretable, which is an attribute that is difficult to accomplish in traditional statistical methods.

1 Introduction

The Internet has rapidly grown into a powerful medium for disseminating information. People can easily share experiences and emotions anytime and anywhere on social media websites. Human feelings can be quickly collected through emotion classification, as these emotions reflect an individual's feelings and experiences toward some subject matters (Turney, 2002; Wilson et al., 2009). Moreover, people can obtain more sponsorship opportunities from manufacturers if their articles about a certain product are able to create more emotional resonance in the readers. Therefore, emotion classification has been attracting more and more attention, e.g., Chen et al. (2010), Purver and Battersby (2012).

Emotion classification aims to predict the emotion categories (e.g., *happy* or *angry*) of the given text (Quan and Ren, 2009; Das and Bandyopadhyay, 2009). There are two aspects of emotions in texts, namely, writer's and reader's emotions. The former concerns the emotion expressed by the writer of the text, and the latter concerns the emotion a reader had after reading it. Recognizing reader-emotion is different and may be even more complex than writer-emotion (Lin et al., 2008; Tang and Chen, 2012). A writer may directly express her emotions through sentiment words. By contrast, reader-emotion possesses a more perplexing nature, as even common words can invoke different types of reader-emotions depending on the reader's personal experiences and knowledge (Lin et al., 2007). For instance, a sentence like "*Kenya survivors describe deadly attack at Garissa University*" is simply stating the facts without any emotion, but may invoke emotions such as *angry* or *worried* in its readers.

In light of this rationale, we propose a principle-based approach (PBA) for reader-emotion classification. It is a highly automated process that integrates various types of knowledge to generate discriminative linguistic templates that can be acknowledged as the essential knowledge for humans to understand different kinds of emotions. PBA recognizes reader-emotions of documents using an alignment algorithm that allows a template to be partially matched through a statistical scoring scheme. Experiments demonstrate that PBA can achieve a better performance than other well-known text categorization methods and the state-of-the-art reader-emotion classification method. Furthermore, we adopt these generated templates to assist in emotional resonance writing. Results show that writers are able to generate more emo-

*Corresponding author

Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers), pages 775–780,
Beijing, China, July 26-31, 2015.

$$-2log\left[\frac{p(w)^{N(w\wedge E)}(1-p(w))^{N(E)-N(w\wedge E)}p(w)^{N(w\wedge\neg E)}(1-p(w))^{N(\neg E)-N(w\wedge\neg E)}}{p(w|E)^{N(w\wedge E)}(1-p(w|E))^{N(E)-N(w\wedge E)}p(w|\neg E)^{N(w\wedge\neg E)}(1-p(w|\neg E))^{N(\neg E)-N(w\wedge\neg E)}}\right] \quad (1)$$

tional resonance in readers after exploiting these templates, demonstrating the capability of PBA in extracting templates with high interpretability.

2 Extracting Emotion Templates from Raw Text

PBA attempts to construct emotion templates through recognition of crucial elements using a three-layered approach. First, since keywords contain important information, PBA learns reader-emotion specific keywords using an effective feature selection method, log likelihood ratio (LLR) (Manning and Schütze, 1999). It employs Equation (1) to calculate the likelihood of the assumption that the occurrence of a word w in reader-emotion E is not random. In (1), E denotes the set of documents of the reader-emotion in the training data; $N(E)$ and $N(\neg E)$ denote the numbers of documents that do or do not contain this emotion, respectively; and $N(w \wedge E)$ is the number of documents with emotion E and having w. The probabilities $p(w)$, $p(w|E)$, and $p(w|\neg E)$ are estimated using maximum likelihood estimation. A larger LLR value is considered closely associated with an emotion. Words in the training data are ranked by LLR values, and the top 200 are included in our emotion keyword list.

Next, named entities (NEs) have been shown to improve the performance of identifying topics (Bashaddadh and Mohd, 2011). Thus, we utilize Wikipedia to semi-automatically label NEs with their semantic classes, which can be considered as a form of generalization. Wikipedia's category tags are used to label NEs recognized by the Stanford NER[1]. If there are more than one category tag for an NE, we select the most dominant one with the highest number of associated Wikipedia pages. The assumption is that the generality of a tag is indicated by the number of Wikipedia pages that are linked to it. For example, a query "奥巴马 (Obama)" to the Wikipedia would return a page titled "贝拉克.奥巴马 (Barack Obama)". Within this page, there are a number of category tags such as "民主党 (Democratic Party)" and "美国总统 (Presidents of the United States)". Suppose "美国总统 (Presidents of the United States)" has more out-going links, we will label "奥巴马 (Obama)" as "[美国总统] (Presidents of the United States)". We also annotate those NEs not found in Wikipedia with their category tags. In this manner, we can transform plain NEs to a more general class and increase the coverage of each label. Finally, to incorporate richer semantic context, we exploit the Extended HowNet (E-HowNet) (Chen et al., 2005) after the above processes to tag the remaining text with sense labels. Figure 1 illustrates crucial element labeling process. Consider the clause C_n = "奥巴马又代表民主党赢得美国总统选举 (Obama, representing the Democratic Party, won the U.S. Presidential election)". First, "奥巴马 (Obama)" is found in the emotion keyword list and tagged. Then, NEs like "民主党 (Democratic Party)" and "总统选举 (Presidential election)" are recognized and tagged as "{政党 (Party)}" and "{总统选举 (Presidential election)}". Subsequently, other terms such as "代表 (represent)" and "赢得 (won)" are labeled with their corresponding E-HowNet senses. Finally, we obtain the sequence "[奥巴马] : {代表} : {政党} : {得到} : {国家} : {总统选举} ([Obama] : {represents} : {party} : {got} : {country} : {Presidential election})." This three-layered labeling process serves as a generalization of the raw text for capturing crucial elements used in the template generation stage that follows.

The emotion template generation process aims at automatically constructing representative templates consisting of a sequence of crucial elements. We observed that the rank-frequency distribution of the elements follows the Zipf's law (Manning and Schütze, 1999). Thus, we rank the templates according to their frequency, and adopt a dominating set algorithm (Johnson, 1974) to use the top 20% templates to cover the rest. First, we constructed a directed graph $G = \{V, E\}$, in which vertices V contains all crucial element sequences $\{CES_1, \cdots, CES_m\}$ in each emotion, and edges E represent the dominating relations between sequences. If CES_x dominates CES_y, there is an edge $CES_x \rightarrow CES_y$. A

[1]http://nlp.stanford.edu/software/CRF-NER.shtml

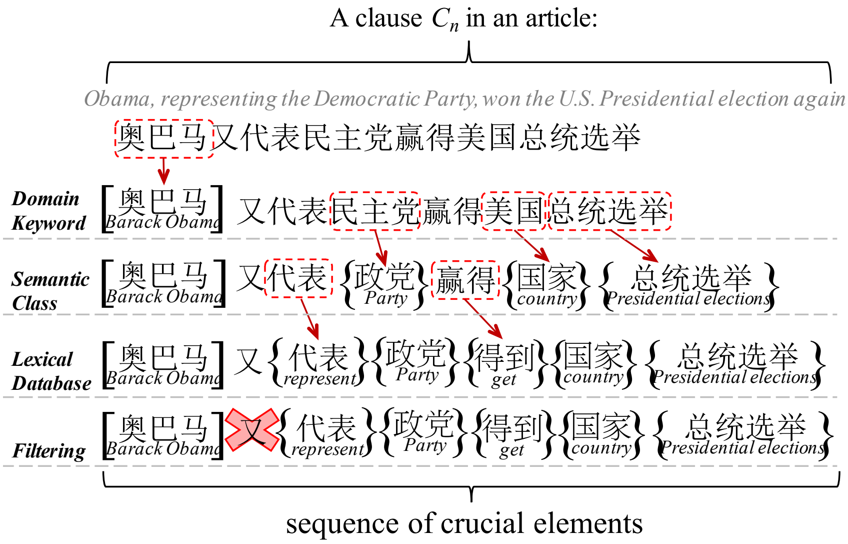

Figure 1: Crucial element labeling process.

dominating relation is defined as follows. 1) Crucial element sequences with high frequency are selected as candidate dominators. 2) Longer sequences dominate shorter ones if their head and tail elements are identical. The intermediate elements are treated as insertions and/or deletions, which can be scored based on their distribution in this emotion during the matching process. Lastly, we preserve top 100 most prominent and distinctive ones from approximately 55,000 sequences based on the dominating rate. This process serves as a kind of dimension reduction and facilitates the execution of our matching algorithm.

3 Template Matching for Inference

We believe that human perception of an emotion is through recognizing important events or semantic contents. For instance, when an article contains strongly correlated words such as "Japan (country)", "Earthquake (disaster)", and "Tsunami (disaster)" simultaneously, it is natural to conclude that this article is more likely to elicit emotions like depressed and worried rather than happy and warm. Following this line of thought, PBA uses an alignment algorithm (Needleman and Wunsch, 1970) to measure the similarity between templates and texts. It enables a single template to match multiple semantically related expressions with appropriate scores. For each clause in a document d_j, we first label crucial elements $CE = \{ce_1, \cdots, ce_n\}$, followed by the matching procedure that compares all sequences in CE from d_j to all emotion templates $ET = \{et_1, \cdots, et_j\}$ in each emotion category, and calculates the scores. Within the alignment matching, a statistical based scoring criterion is used to score insertions, deletions, and substitutions as described below. The emotion e_i with the highest sum of scores defined in (2) is considered as the winner.

$$
\begin{aligned}
&Emotion(d_j) \\
&= \arg\max_{e_i \in E} \sum_{et_n \in ET_{c_i}, ce_m \in CE_{d_j}} \Delta(et_n, ce_m) \\
&= \sum_k \sum_l \Delta(et_n \cdot sl_k, ce_m \cdot ce_l)
\end{aligned} \tag{2}
$$

where sl_k and ce_l represent the k^{th} slot of et_n and l^{th} element of ce_m, respectively. Details for scoring matched and unmatched elements are as follows. If $et_n \cdot sl_k$ and $ce_m \cdot ce_l$ are identical, we add a matched score (*MS*) obtained from the LLR value of ce_l if it matches a keyword. Otherwise, the score is determined by multiplying the frequency of the crucial element in category c_i by a normalizing factor $\lambda = 100$, as in (3). On the other hand, an unmatched element is given a score of insertion or deletion. The insertion score (*IS*), defined as (4), can be accounted for by the inversed entropy of this element, which represents the uniqueness or generality of it among categories. And the deletion score (*DS*), defined as (5), is computed from the log frequency of this crucial element in this emotion category.

$$MS(ce_l) = \begin{cases} LLR_{ce_l}, \text{if } ce_l \in \text{keyword} \\ \lambda \frac{f_{ce_l}}{\sum_{i=1}^{m} f_{ce_i}}, \text{otherwise} \end{cases} \quad (3)$$

$$IS(ce_l) = \frac{1}{-\sum_{i=1}^{m} P(ce_{l_{c_i}}) \cdot log_2(P(ce_{l_{c_i}}))} \quad (4)$$

$$DS(ce_l) = log \frac{f_{ce_l}}{\sum_{i=1}^{m} f_{ce_i}} \quad (5)$$

4 Experiments

4.1 Dataset and Setting

We collected a corpus of Chinese news articles from Yahoo! Kimo News[2], in which each article is given votes from readers with emotion tags in eight categories: *angry*, *worried*, *boring*, *happy*, *odd*, *depressing*, *warm*, and *informative*. We consider the voted emotions as the reader's emotion toward the news. Following previous studies such as Lin et al. (2007) and Lin et al. (2008) that used a similar source, we exclude "*informative*" as it is not considered as an emotion category. To ensure the quality of our evaluation, only articles with a clear statistical distinction between the highest vote of emotion and others determined by *t*-test with a 95% confidence level are retained. Finally, 47,285 out of 68,026 articles are kept, and divided into the training set and the test set, each containing 11,681 and 35,604 articles, respectively.

4.2 Reader's Emotion Classification

Several classification methods are implemented and compared. Naïve Bayes (McCallum et al., 1998) serves as the baseline, denoted as *NB*. In addition, a probabilistic graphical model that uses LDA as document representation to train an SVM classifier that determines a document as either relevant or irrelevant (Blei et al., 2003), denotes as *LDA*. Next, an emotion keyword-based model, denoted as *KW*, is trained using SVM to test the effect of our keyword extraction approach. *CF* is the state-of-the-art reader-emotion recognition method that combines various features including bigrams, words, metadata, and emotion category words (Lin et al., 2007). For evaluation, we adopt the accuracy measures as used by Lin et al. (2007), and compute the macro-average (A^M) and micro-average (A^μ). Table 1 shows a comprehensive evaluation of PBA and other methods[3].

Emotion	*Accuracy(%)*				
	NB	**LDA**	**KW**	**CF**	**PBA**
angry	47.00	74.21	79.21	83.71	**87.83**
worried	69.56	**92.83**	81.96	87.50	75.80
boring	75.67	76.21	84.34	87.52	**90.52**
happy	37.90	67.59	80.97	86.27	**88.94**
odd	73.90	**85.40**	77.05	84.25	83.34
depressing	73.76	81.43	85.00	87.70	**92.15**
warm	75.09	87.09	79.59	85.83	**91.91**
A^M	58.11	76.10	80.36	85.80	**86.43**
A^μ	52.78	74.16	80.81	85.70	**88.56**

Table 1: Comparison of the accuracies of five reader-emotion classification systems.

Since *NB* only considers surface word weightings and ignore inter-word relations, it only achieved an accuracy of 58.11%. By contrast, *LDA* includes both keywords and long-distance relations, thus greatly outperforms *NB* with an overall accuracy of 76.10%. It even obtained the highest accuracy of 92.83% for the emotions "*worried*" and "*odd*" among all methods. Notably, *KW* can bring about substantial proficiency in detecting the emotions, which indicates that reader-emotion can be recognized effectively by using only the LLR scores of keywords. Meanwhile, *CF* achieved a satisfactory overall accuracy around 85%, due to the combined lexical feature sets (e.g., character bigrams, word dictionary, and emotion keywords), paired with metadata of the articles. For instance, we found that many sports-related articles invoke the emotion "*happy*". Specifically, 45% of all "*happy*" instances are sports-related, and a sports-related article has a 31% chance of having the emotion tag "*happy*". Hence, the high accuracy of the emotion category "*happy*" could be the result of people's general enthusiasm over sports rather than a particular event. On top of that, PBA can generate distinctive emotion templates to capture variations of similar expressions, thus achieving better outcome. For instance, the template "{国家 $_{\text{country}}$}"：[发生 $_{\text{occur}}$]：[地震 $_{\text{earthquake}}$]：{劫难 $_{\text{disaster}}$}" is generated by PBA

[2]https://tw.news.yahoo.com

[3]The dictionary required by all methods is constructed by removing stop words according to (Zou et al., 2006), and retaining tokens that make up 90% of the accumulated frequency. As for unseen events, we used the Laplace smoothing in *NB*. *LDA* is implemented using a toolkit (http://nlp.stanford.edu/software/tmt/tmt-0.4/).

for the emotion "*worried*". It is perceivable that this template is relaying information about disastrous earthquakes in a country, and such news often makes readers worry. The ability to yield such emotion-specific, human interpretable templates could account for the outstanding performance of PBA.

4.3 Reader's Emotion Templates Suggestion in Emotional Resonance Writing

This experiment aims at testing the effectiveness of the emotion templates in aiding writers to compose articles with stronger emotional resonance. Here, we only consider coarse-grained emotion categories (i.e., *positive* and *negative*). Thus, fine-grained emotions like *happy*, *warm*, and *odd* are merged into '*positive*', while *angry*, *boring*, *depressing*, and *worried* are merged into '*negative*'. 10 templates for each of the fine-grained emotions are selected, resulting in 30 and 40 templates for the two coarse-grained emotions, respectively. We recruited seven writers to compose two articles that they think will trigger positive and negative emotions without using templates (denoted as *NT*). Then, we asked them to compose two more articles with the aid of templates (denoted as *WT*). Afterwards, all articles are randomly organized into a questionnaire to test the emotional resonance. Subjects are required to perform two tasks: 1) answer 'positive', 'neutral', or 'negative' 2) give a score according to the five-point Likert scale (Likert, 1932) for a given emotion. In the end, 42 effective responses are gathered. For Task 1, the score is defined as the number of matching responses and answers. As for Task 2, the score is the sum of all articles. Higher scores indicate better emotional resonance between writers and readers. Figure 2 shows the sum of scores in Task 1 from all subjects, grouped by writer. As for Task 2, Figure 3 shows the average score across subjects, grouped by writers. In both tasks, we can see that higher scores are obtained after using templates, indicating that emotion templates can indeed assist writers in creating stronger emotional resonance in their composition.

To sum up, results show that PBA can generate emotion templates that not only help machines predict reader's emotion, but also effectively aid writers in creating a stronger emotional resonance with the readers.

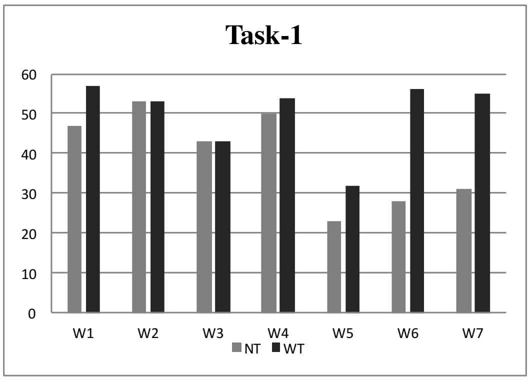

Figure 2: Comparison of the number of correct emotional response before and after utilizing templates.

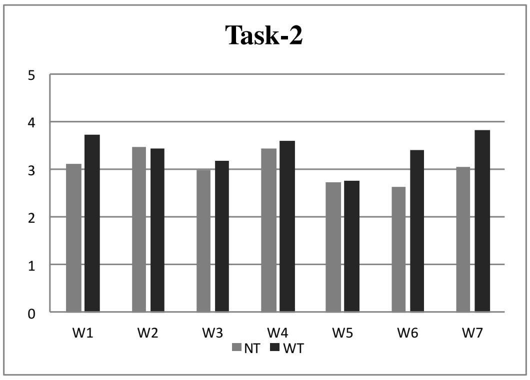

Figure 3: Degree of emotional resonance between writers and readers.

5 Conclusion

In this paper, we present PBA, a flexible, highly automated, and human-interpretable approach for reader-emotion classification. By capturing prominent and representative patterns in texts, PBA can effectively recognize the reader-emotion. Results demonstrate that PBA outperforms other reader-emotion detection methods, and can assist writers in creating higher emotional resonance. In the future, we plan to further refine and employ it to other NLP applications. Also, additional work can be done on combining statistical models into different components of PBA.

Acknowledgments

We are grateful to the anonymous reviewers for their insightful comments. This research was supported by the Ministry of Science and Technology of Taiwan under grant MOST 103-3111-Y-001-027, 103-2221-E-002-106-MY2, and NSC102-3113-P-001-006.

References

Omar Mabrook A Bashaddadh and Masnizah Mohd. 2011. Topic detection and tracking interface with named entities approach. In *Proceedings of the International Conference on Semantic Technology and Information Retrieval (STAIR)*, pages 215–219.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Keh-Jiann Chen, Shu-Ling Huang, Yueh-Yin Shih, and Yi-Jun Chen. 2005. Extended-HowNet - a representational framework for concepts. In *Proceedings of OntoLex 2005 - Ontologies and Lexical Resources IJCNLP-05 Workshop*.

Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 179–187.

Dipankar Das and Sivaji Bandyopadhyay. 2009. Word to sentence level emotion tagging for bengali blogs. In *Proceedings of the ACL-IJCNLP 2009 Conference*, pages 149–152.

David S. Johnson. 1974. Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, 9(3):256 – 278.

Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 140:1–55.

Kevin Hsin-Yih Lin, Changhua Yang, and Hsin-Hsi Chen. 2007. What emotions do news articles trigger in their readers? In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 733–734.

Kevin Hsin-Yih Lin, Changhua Yang, and Hsin-Hsi Chen. 2008. Emotion classification of online news articles from the reader's perspective. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, volume 1, pages 220–226.

Christopher D Manning and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. MIT press.

Andrew McCallum, Kamal Nigam, et al. 1998. A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization*, volume 752, pages 41–48.

Saul B Needleman and Christian D Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453.

Matthew Purver and Stuart Battersby. 2012. Experimenting with distant supervision for emotion classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 482–491.

Changqin Quan and Fuji Ren. 2009. Construction of a blog emotion corpus for chinese emotional expression analysis. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, volume 3, pages 1446–1454.

Yi-jie Tang and Hsin-Hsi Chen. 2012. Mining sentiment words from microblogs for predicting writer-reader emotion transition. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*, pages 1226–1229.

Peter D Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3):399–433.

Feng Zou, Fu Lee Wang, Xiaotie Deng, Song Han, and Lu Sheng Wang. 2006. Automatic construction of chinese stop word list. In *Proceedings of the 5th WSEAS International Conference on Applied Computer Science*, pages 1010–1015.