

**High-Performance Computing Account**  
**Application Supplementary Report:**  
**Reader-Emotion Classification of News Articles**

Hsin-Yih Lin

Changhua Yang

Hsin-Hsi Chen

Department of Computer Science and Information Engineering  
National Taiwan University

## **1. INTRODUCTION**

Past researches on emotions conveyed by documents focused on detecting the feelings that the authors of the documents were expressing. Such studies are useful when we need a quick poll on how people feel about a particular event or item.

Instead of uncovering the emotional states of the authors of documents, we aim to find out what emotions documents trigger in their readers. Such research has novel applications such as incorporating emotion search information retrieval.

## **2. CORPUS**

We used Yahoo!'s Chinese news articles as our corpus. Yahoo!'s Chinese news webpage has a special feature which allows a reader of a news article to express how he or she feels after reading the news article. The eight choices that a reader may select from to describe feeling are *happy*, *angry*, *sad*, *surprised*, *heartwarming*, *awesome*, *bored* and *useful*.

Our corpus contains news articles spanning a period of 81 days. Articles that were published during the first 54 days and those that were published in the last 27 days were used as training and testing data, respectively. The training data contained 12,079 articles and the testing data contained 5,664 articles. For each news article, we were able to obtain the number of votes of each emotion. We collected an article seven days after it was published to ensure that the votes have stabilized so that the rank of each emotion would no longer be subject to change.

## **3. CLASSIFICATION METHOD**

Our goal was to classify the news articles into one of the 8 emotion classes provided by Yahoo! Chinese news. Several features were extracted from the news

articles. The first set of features consists of all the Chinese character bigrams that appear in the articles. For the second set of features, we applied Stanford NLP group’s Chinese segmentation tool on the title and content of each article. The words output by the segmentation tool were used. The third set is the metadata of the articles, which are the news reporter, news category, location of the news event, time (hour of publication) and news agency. News category refers to the newspaper sections such as business and political. The fourth set is the emotion categories of words. Yahoo! Kimo Blog service provides users 40 emotion categories. Words can be assigned emotion categories and weights by machine learning using the Yahoo! Kimo Blog data sets. LibSVM was used as the classifier. We used binary weights for the first three sets of features. The last set of features had weights ranging from 0 to 1, depending on the relative frequencies of each of the 40 emotion categories.

#### 4. EXPERIMENTS AND DISCUSSION

**Table 1. Testing accuracy using different feature combinations**

BI	WD	MT	BI+WD	BI+MT	WD+MT	BI+WD+MT	BI+WD+MT+EC
70.81%	69.75%	49.77%	71.11%	71.85%	70.51%	72.01%	72.05%

**Table 2. Percentage of correctly classified testing instances for each emotion class**

Features	<i>Awesome</i>	<i>Heart-warming</i>	<i>Surprised</i>	<i>Sad</i>	<i>Useful</i>	<i>Happy</i>	<i>Bored</i>	<i>Angry</i>
BI+WD+MT	62.10%	62.40%	58.81%	64.40%	89.81%	77.44%	79.87%	74.49%

For the experiments, we tried different combinations of features. The results are shown in Table 1. BI, WD, MT and EC denote bigram, word, metadata, and emotion category of word, respectively. Since news article readers rarely vote unanimously for

a single emotion class, votes are usually distributed among several emotions in an article. We chose the top-ranking emotion class of each instance as the correct class.

Table 1 shows that BI performs better than WD, and BI+MT performs better than WD+MT. Holding all other conditions the same, Chinese character bigrams are better features than the segmented words. However, using BI and WD in combination with MT produces better accuracies than using BI and WD separately.

In Table 1, BI+WD+MT+EC performs almost as well as BI+WD+MT in all rows. So, the emotion category of words has certain influence on the classification accuracy.

Table 2 shows that different classes have very different performance, ranging from having 58.81% correctly-classified instances to 89.81% using features BI+WD+MT. We were concerned that having highly distinguishing event words as features may be the cause of *happy* and *angry* having high percentage figures relative to other emotion classes. As event words may occur only for a short period of time and rarely be used again in future news stories, having event words as the primary distinguishing features is not going to be helpful in enhancing the general coverage of the classification system. To find out if this was really the case, we examined the most frequently occurring features for each class and computed the conditional probability  $P(\text{instance } i\text{'s true class is } c | \text{instance } i \text{ has feature } f)$  to give an indication of how distinguishing these features were. For *happy* class, we found that a feature shared by many instances is the news category sports. In particular, 48% of all *happy* instances belong to the news category sports. It is also observed that an instance with the news category sports has 67% chance of having the true class *happy*. So, the high accuracy of *happy* class can be a result of people's general enthusiasm over sports rather than a result of a particular event.

A similar observation was obtained for *bored* class. Features that were shared by a great number of *bored* instances were mostly general terms related to politics

instead of a particular time-dependent event. These features included the Chinese character bigrams that denote “legislator” and “minister”. One of the most shared features is the political news category metadata. Any training instance in the political category has 64% chance of belonging to *bored* class.

In Table 2, *useful* class has the highest percentage of correctly classified testing instances. As we examined the most frequently occurring features of *useful* class, we found a lot of features related to weather. An instance with the weather category has 84% chance of belonging to *useful* class in the training set and 95% chance of belonging to *useful* class in the testing set. The weather forecast news articles contribute to the performance of *useful* class.

## **5. CONCLUSION**

The combination of bigrams, words, metadata and word emotion categories achieves the best accuracy in readers’ emotion classification, i.e., 72.05%. We will study additional cues to extend the methodology to other types of documents and ultimately integrate it into IR systems.