Multi-label Text Classification Approach for Sentence Level News Emotion Analysis

Plaban Kr. Bhowmick, Anupam Basu, Pabitra Mitra, and Abhishek Prasad

Indian Institute of Technology, Kharagpur, India-721302

Abstract. Multiple emotions are often evoked in readers in response to text stimuli like news article. In this paper, we present a novel method for classifying news sentences into multiple emotion categories using Multi-Label K Nearest Neighbor classification technique. The emotion data consists of 1305 news sentences and the emotion classes considered are disgust, fear, happiness and sadness. Words and polarity of subject, verb and object of the sentences and semantic frames have been used as features. Experiments have been performed on feature comparison and feature selection.

1 Introduction

The Internet has been one of the primary media for information dissemination shortly after the advent of Word Wide Web. Consequently, the amount of text data available online is enormous. The advent of new technologies makes way of new interaction possibilities and provides people with an array of social media like blog, chat, social network, news etc. As compared to traditional keyword based or topical access to the information, new social interactions require the information to be analyzed in social dimensions like emotion, sentiment, attitude, belief etc. Emotion analysis of text is one of such tasks that are gaining importance in text mining community in recent times.

Two types of emotion evaluation may be possible: writer perspective and reader perspective evaluation. In previous works [1], a few attempts towards writer perspective analysis of emotion in text data have been made. In all these studies, it has generally been assumed that the writer expresses only one emotion for a text segment. However, evocation of a blend of emotions is common in reader in response to a stimulus. For example, the following sentence may evoke *fear* and *sad* emotion in readers mind.

`Militant attack kills over 30 persons in Nigeria.`

Classifying emotion from reader's perspective is a challenging task and research on this topic is relatively sparse as compared to writer perspective analysis.

Affective text analysis was the task set in *SemEval-2007 Task 14* [2]. A corpus of news headlines extracted from Google news and CNN was provided. Some systems with different approaches have participated in solving the said task [2].

The work [3] provides the method for ranking reader's emotions in Chinese news articles from Yahoo! Kimo News. Eight emotional classes are considered

S. Chaudhury et al. (Eds.): PReMI 2009, LNCS 5909, pp. 261–266, 2009.

in this work. Support Vector Machine (SVM) has been used as the classifier. Chinese character bigram, Chinese words, news metadata, affix similarity and word emotion have been used as features. The best reported system accuracy is 76.88%.

In this work, we perform reader perspective emotion analysis in text data where one text segment may evoke more than one emotion in reader. News is a media where certain facts in the articles are presented to the readers with the expectation that the articles evoke some emotional responses in the readers. So, this media is one potential data source for the computational study of reader perspective emotion.

2 Emotion Classification Problem and MLκNN

The problem of multi-label emotion classification is defined as follows: Let $S = \{s_1, s_2, \ldots, s_n\}$ be the set of emotional sentences and $\mathcal{E} = \{e_i | i = 1, 2, \ldots, |\mathcal{E}|\}$ be the set of emotion classes (e.g., happy, sad etc.). The task is to find a function $h : S \mapsto 2^{\mathcal{E}}$, where $2^{\mathcal{E}}$ is the powerset of $\mathcal{E}$.

This problem can be mapped to a multi-label text classification problem. In this work, we use Multi-Label κ Nearest Neighbor classifier (MLκNN) [4] for classifying sentences into emotion classes. MLκNN, a multi-label adaptation of single label κ Nearest Neighbor algorithm, is one of the state of the art high performance algorithm adaptation technique. In this technique, for each test instance t, its K nearest neighbors in the training set are identified. Then according to statistical information gained from the label sets of these neighboring instances maximum a posteriori (MAP) principle is utilized to determine the label set for the test instance. The entities that central to this classification technique are the prior probabilities $P(H_b^l)(l \in \mathcal{E}),\, b \in \{0, 1\})$ and the posterior probabilities $P(N_j^l | H_b^l)\,(j \in \{0, 1, \ldots, \kappa\})$. Here, H_1^l is the event that the test instance has label l, while H_0^l denotes the event that t has not label and $N_j^l\,(j = 1, 2, \ldots, \kappa\,)$ is denotes that among the κ nearest neighbors of t, there are exactly j instances which have label l. The probability values are estimated from training data set. Laplace smoothing is used for handling data sparsity problem.

3 Emotion Data

The emotion text corpus consists of 1305 sentences extracted from *Times of India* news paper archive[1]. The emotion label set consists of four emotions: disgust, fear, happiness and sadness. A sentence may trigger multiple emotions simultaneously. So, one annotator may classify a sentence into more than one emotion categories.

The distribution of sentences across emotion categories is as follows: Disgust = 307, Fear = 371, Happiness = 282 and Sadness = 735. The *label density* (LD $= \frac{1}{|S|}\sum_{i=1}^{|S|}\frac{|\mathcal{E}_i|}{|\mathcal{E}|}$, where $\mathcal{E}_i$ is the label set for S_i) and *label cardinality*

[1] http://timesofindia.indiatimes.com/archive.cms

($\mathrm{LC} = \frac{1}{|S|}\sum_{i=1}^{|S|}|\mathcal{E}_i|$) are the measures of how multi-label the data is. The LC and LD for this data set are computed to be 1.3 and 0.26 respectively.

4 Features for Emotion Classification

Three types of features have been considered in our work as given below:

- *Word Feature (W)*: Words sometimes are indicative of the emotion class of a text segment. For example, the word 'bomb' may be highly co-associated with fear emotion. Thus words present in the sentences are considered as features. Before creating the word feature vectors, stop words and named entities are removed and the words are stemmed using Porter's stemmer.
- *Polarity Feature (P)*: Polarity of the subject, object and verb of a sentence may be good indicators of the emotions evoked. The subject, object and verb of a sentence is extracted from its parse tree and the polarity for each phrase is extracted from manual word level polarity tagging with a set of simple rules.
- *Semantic Frame Feature (SF)*: The Berkeley FrameNet project[2] is a well-known resource of frame-semantic lexicon for English. Apart from storing the predicate-argument structure, the frames group the lexical units. For example, the terms 'kill', 'assassin' and 'murder' are grouped into a single semantic frame 'Killing'. In this work, we shall be exploring the effectiveness of the semantic frames feature in emotion classification. The semantic frame assignment was performed by SHALMANESER[3].

5 Evaluation Measures

We evaluate our emotion classification task with respect to different sets of multi-label evaluation measures:

- Example based measures: Hamming Loss (HL), Partial match accuracy (P-Acc), Subset accuracy (S-Acc) and F1. These measures are explained in the work [5].
- Ranking based measures: One Error (OE), Coverage (COV), Average Precision (AVP). The work [4] describes these measures in detail.

6 Experimental Results

In this section, we present results of experiments of emotion classification with MLkNN. 5-fold cross-validation has been performed in all the experiments and the number of neighbors considered is 10.

[2] http://framenet.icsi.berkeley.edu/

[3] http://www.coli.uni-saarland.de/projects/salsa/shal/

6.1 Comparison of Features

The comparison of the features is performed with respect to a baseline which considers only words as features. Table 1 summarizes the results of emotion classification with different features and their combinations with best results presented in bold face.

Table 1. Comparison of features (W = word feature, P = polarity feature, SF = Semantic frame feature)

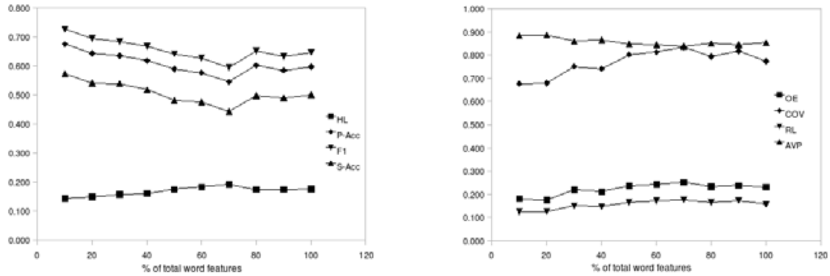
Measure Type	Measure	W	P	SF	W+P	P+SF	W+SF	W+P+SF
Example based measures	HL	0.175	0.223	**0.117**	0.156	0.118	0.131	0.126
	P-Acc	0.598	0.532	0.756	0.664	**0.764**	0.703	0.722
	F1	0.647	0.582	0.806	0.714	**0.817**	0.756	0.774
	S-Acc	0.499	0.428	0.653	0.563	**0.655**	0.594	0.618
Ranking based measures	OE	0.231	0.325	0.143	0.210	**0.129**	0.145	0.133
	COV	0.774	0.918	0.576	0.704	**0.538**	0.606	0.576
	RL	0.157	0.207	0.091	0.138	**0.079**	0.101	0.090
	AVP	0.853	0.801	0.911	0.871	**0.921**	0.906	0.915

General observations over the feature comparison experiment are as follows.

- The use of semantic frames (**SF**) as features improves the performance of emotion classification significantly ($\triangle$P-Acc = 15.8%, $\triangle$S-Acc = 15.4% and $\triangle$F1 = 15.9%) over the baseline. This significant improvement may be attributed to two different transformations over the word feature set.
  - *Dimensionality Reduction*: There is a significant reduction in dimension when (**SF**) feature is considered instead of (**W**) feature (SF feature dimension = 279 and W feature dimension = 2345).
  - *Feature Generalization*: Semantic frame assignment to the terms in the sentences is one of generalization technique where conceptually similar terms are grouped into a semantic frame. In semantic frame feature set, unification of these features are performed resulting in less skewedness in feature distribution.
- The **P+SF** feature combination performs best.
- The polarity feature (**P**) is inefficient as compared to other combinations but whenever coupled with other feature combinations (i.e., **W** vs. **W+P**, **SF** vs. **SF+P** and **W+SF** vs. **W+SF+P**), the performance improves.
- Whenever **W** feature combines with **SF**, degradation in performance have been observed (i.e., **SF** vs. **W+SF**, **P+SF** vs. **W+P+SF**).

6.2 Feature Selection

The plot word feature χ^2 value vs. rank follows the Zipfian distribution (power law fit with equation $y = \alpha x^{-\beta}$ where $\alpha = 236.43$, $\beta = 0.8204; R^2 = 0.89$) having a long tail which is strong indication of feature sparseness problem. To



(a) Example based measures Vs. Percentage of total word features

(b) Ranking based measures Vs. Percentage of total word features

Fig. 1. Variation of multi-label measures with Percentage of total word features

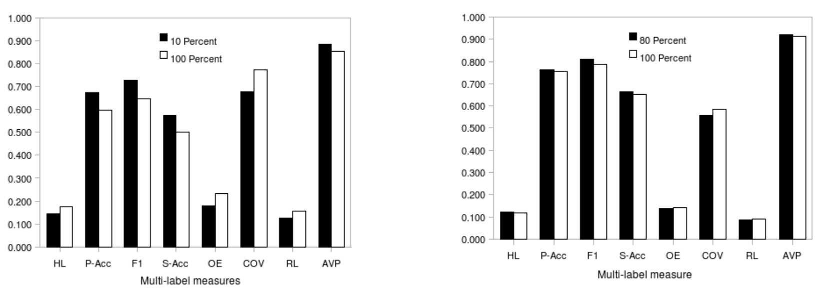
alleviate this problem, we have performed χ^2 feature selection [6] on the **W** and **SF** feature sets.

We performed experiment on selecting optimal **W** feature set size based on their χ^2 values. We present the variations of the performance measures with **W** feature set size in Fig. 1.

Top 10% of the total **W** feature set is found to be optimal feature set. The relative performance after feature selection for **W** is shown in Fig. 2(a). In case of **SF** feature, top 80% out of the total set was selected as optimal feature set for **SF** feature. The relative performance with the selected **SF** feature set is presented in Fig 2(b).

It is evident from Fig. 2 that the there is a slight improvement in performance after adopting feature selection strategy for both the feature sets.

With **P+SF** feature combination being the close competitor, best performance is achieved with **P+80%SF**(HL = 0.115, P-Acc = 0.773, F1 = 0.827 and S-Acc = 0.666). As the emotion analysis task is modeled in a multi-label framework, comparison with other systems can only be made with micro-average



(a) Relative performance after χ^2 word feature selection

(b) Relative performance after χ^2 semantic frame feature selection

Fig. 2. Performance after χ^2 feature selection

Table 2. Comparison of the proposed system with others

Measure	Accuracy	Precision	Recall	F1
UPAR7	89.43	27.56	5.69	9.43
UA-ZBSA	85.72	17.83	11.27	13.81
SWAT	88.58	19.46	8.62	11.95
Li and Chen	76.88	–	–	–
Our System	**88.2**	**84.42**	**79.93**	**82.1**

measures like accuracy, precision, recall and F1. The comparison with other systems is presented in Table 2.

7 Conclusions

We have presented an extensive comparison of different features for multi-label reader perspective emotion classification with MLkNN. Feature selection on word and semantic frame features is found to be fruitful for better performance. The classifier performs better with semantic frame (SF) features compared to word features (W) as SF helps in dealing with feature sparseness problem. Improvements have been noticed when the polarity feature is combined with other features.

References

1. Abbasi, A., Chen, H., Thoms, S., Fu, T.: Affect analysis of web forums and blogs using correlation ensembles. IEEE Transactions on Knowledge and Data Engineering 20(9), 1168–1180 (2008)
2. Strapparava, C., Mihalcea, R.: Semeval-2007 task 14: Affective text. In: Proceedings of the 4th International Workshop on the Semantic Evaluations (SemEval 2007), Prague, Czech Republic (June 2007)
3. Lin, K.H.Y., Yang, C., Chen, H.H.: Emotion classification of online news articles from the reader's perspective. In: Web Intelligence, pp. 220–226 (2008)
4. Zhang, M., Zhou, Z.h.: Ml-knn: A lazy learning approach to multi-label learning. Pattern Recognition 40, 2007 (2007)
5. Tsoumakas, G., Vlahavas, I.: Random *k*-labelsets: An ensemble method for multi-label classification. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) ECML 2007. LNCS (LNAI), vol. 4701, pp. 406–417. Springer, Heidelberg (2007)
6. Zheng, Z., Wu, X., Srihari, R.: Feature selection for text categorization on imbalanced data. Explorations Newsletter 6(1), 80–89 (2004)