Reader Perspective Emotion Analysis in Text through Ensemble based Multi-Label Classification Framework

Plaban Kumar Bhowmick (Corresponding author)

Indian Institute of Technology Kharagpur

Kharagpur, India - 721302

E-mail: plaban@gmail.com

Anupam Basu

Indian Institute of Technology Kharagpur

Kharagpur, India - 721302

E-mail: anupam@iitkgp.ac.in

Pabitra Mitra

Indian Institute of Technology Kharagpur

Kharagpur, India - 721302

E-mail: pabitra@cse.iitkgp.ernet.in

Abstract

Multiple emotions are often triggered 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 an ensemble based multi-label classification technique called RAKEL. The emotion data consists of 1305 news sentences and the emotion classes considered are disgust, fear, happiness and sadness. Words are the most obvious choice as feature for emotion recognition. In addition to that we have introduced two novel feature sets: polarity of subject, verb and object of the sentences and semantic frames. Experiments concerning the comparison of features revealed that semantic frame feature combined with polarity based feature performs best in emotion classification. Experiments on feature selection over word and semantic frame features have been performed in order to handle feature sparseness problem. In both word and semantic frame feature, improvements in the overall performance have been observed after optimal feature selection.

Keywords: Emotion classification, Multi-label classification, Ensemble classifier, Feature selection

1. Introduction

In the area of Natural Language Processing (NLP), syntactic and semantic level processing of text has been the focus of attention for decades. The related tasks like parts of speech tagging, parsing, machine translation, semantic role labeling have been solved to an acceptable accuracy for different languages. With syntactic and semantic tools in the disposal, the NLP researchers are looking forward to solve the challenges that deal with social and humanistic dimensions of text like emotion, sentiment, attitude, belief etc.

Analysis of the views of the users towards a particular entity is the focus of study in Opinion mining or sentiment analysis (B. Pang & L. Lee., 2004). This task judges an entity in the dimension of positivity or negativity, i.e., whether a particular product is liked by the users or not. On the other hand, emotion analysis of text goes beyond positive-negative dimension to discrete emotion categories like happiness, sadness etc.

Facial or audio expressions are the most notable and prominent clues and have widely been used in analyzing emotion. Though emotion is not a linguistic entity (Z. Kovecses., 2003), in many situations, emotion is expressed through language in day-to-day speech communications or published communications.

Emotion can be analyzed from two different perspectives: From the writer/speaker perspective, where we need to understand the emotional intent of the writer/speaker and from the reader's perspective, where we try to identify the emotion that is evoked in a reader in response to a language stimulus. In the current study, we aim at performing sentence level emotion analysis from a reader's perspective which includes the following challenges.

- *Triggering of multiple emotions*: Given a sentence, a mix of multiple emotions can be triggered in a reader. For example, the following sentence may evoke *fear* and *sad* emotion in readers mind.

A 23-year-old pregnant woman succumbed to swine flu at a city-based hospital.

- *Study of suitable features*: Emotion analysis of text being in its infancy, appropriate feature set required for emotion analysis has not been investigated properly.

- *Feature sparseness*: While emotion analysis in discourse or paragraph level may provide larger number of cues as features, in a single sentence, the number of features is less indicating a feature sparseness problem.

Selection of data source is an important issue. We have considered an age old popular concept in news media for writing emotionally charged news articles called *Emotional framing* (P. E. Corcoran., 2006). According to this theory, each news item is shaped into a form of story with layered dramatic frames, e.g., fear caused by danger; sorrow and grief arising from violence, crime and death; exhilaration and joy resulting from good luck or victory. As a result, amount of news articles capable of evoking emotions in readers is huge. Accordingly, we have rested our study on a set of sentences collected from news articles and headlines.

The contributions of this work are as follows:

- *Multi-Label model of emotion*: The problem reader perspective emotion analysis has been modeled in a multi-label classification framework where the sentence has belongingness in multiple emotion categories. Consequently, the problem of reader emotion classification in text data can be mapped to a multi-label text categorization problem. In this work, we use an ensemble based method, RAKEL (Tsoumakas, 2007), for emotion classification.

- *Feature space exploration*: A thorough exploration of features for reader emotion analysis has been performed in the work. Word feature and word co-occurrence statistics have been used in the earlier works towards reader perspective emotion analysis. In addition to word feature, we have introduced two new features, namely the polarity feature (subject, object and verb) and the semantic frame feature. In the baseline study, word occurrence feature based model is considered. The description and extraction methods for these features have been provided. Semantic frames are generalization of terms or words in the lexicon. Use of semantic frames as feature provides the facility of dimensionality reduction and feature generalization. Thus, semantic frame based emotion recognition model outperforms other feature group based models by a considerable margin. Semantic frame feature, coupled with polarity feature performs best in the selected multi-label classification framework.

- *Feature selection study*: Selection of appropriate features is important as it will help in filtering out the redundant and noisy features from the feature space. Feature selection experiments (χ^2 feature selection) on word and semantic frame features have been performed to train the classifier with optimal feature set.

The rest of the paper is organized as follows: In section 2, we review some of the previous works in writer and as well as reader perspective emotion analysis. In section 3, we point out the limitations in the previous works. A formal representation of the multi-label emotion classification problem and a brief description of the multi-label classification framework used in this study have been provided in section 4. We description and statistics of the emotion data set has been presented in section 5. The features used in this study have been provided in section 6. In section 7, we discuss the experimental set up and present the outcomes of different experiments.

2. Related Works

As stated earlier, emotion analysis can be performed in two different perspectives. So, we provide overview of previous works on emotion analysis from both the perspectives.

2.1 Emotion Analysis in Reader Perspective

Affective text analysis was the task set in *SemEval-2007 Task 14* (C. Strapparava & R. Mihalcea., 2007). A corpus of news headlines collected from Google news and CNN was considered in this task. Two types of tasks were considered: To classify headlines into positive/negative emotion category and as well as distinct emotion categories like anger, disgust, fear, happiness, sadness and surprise.

The system UA-ZBSA (Z. Kozareva, B. Navarro, S. Vazquez & A. Montoyo., 2007) computes statistics from three different search engines (MyWay, AllWeb and Yahoo) to label the news headlines with emotion classes. The work derives the PMI score of each content word of a headline with respect to each emotion by querying the search engines

with the headline and the emotion. The accuracy, precision and recall of the system are reported to be 85.72%, 17.83% and 11.27% respectively.

UPAR7 (F. R. Chaumartin., 2007) adopt a rule-based approach towards emotion classification. The system performs emotion analysis on news headline data provided in SemEval-2007 Task 14. The common words are decapitalized with the help of parts of speech tagger and Wordnet (C. Fellbaum., 1998) in the preprocessing step. Each word is first rated with respect to emotion classes. The main theme word,which is detected by parsing a headline, is given a higher weight than the other words. The emotion score boosting to the nouns are performed based on their belongingness to some general categories in Wordnet. The word scoring also considers some other factors like human will, negation and modals, high-tech names, celebrities etc. The average accuracy, precision and recall of the system are 89.43%, 27.56% and 5.69% respectively.

A supervised approach has been adopted by the system SWAT (P. Katz, M. Singleton & R. Wicentowski., 2007) towards emotion classification in news headlines. The system develops a word-emotion map by querying the Roget's New Millennium Thesaurus. The score each word in the headline is assigned with the help of the created map. The average score of the headline words are considered while labeling it with a particular emotion. The reported classification accuracy, precision and recall are 88.58%, 19.46% and 8.62% respectively.

The work by Lin and Chen (K. H. Y. Lin, C. Yang & H. H. Chen., 2008a; K. H. Y. Lin, & H. H. Chen. 2008b) deals with the method for ranking reader's emotions in Chinese news articles from Yahoo! Kimo News. Eight emotional classes are considered in this work. Support vector machine 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%.

2.2 Emotion Analysis in Writer Perspective

Subasic and Huettner (P. Subasic & A. Huettner. 2001) proposed Fuzzy Semantic Typing approach where manually developed fuzzy lexicon was used. In this lexicon, one word may belong to multiple emotion categories with varying intensity and membership values. Many other works (Y. H. Cho & K. J. Lee., 2006) make use of emotion lexical resources for writer perspective emotion analysis. Wordnet-Affect (A. Valitutti, C. Strapparava & O. Stock., 2004) is used in emotion detection task (C. Ma, H. Prendinger & M. Ishizuka., 2005).

Mishne (G. Mishne., 2005) performs emotion analysis on blogpost corpus. The feature set considered in this task consists of frequency counts, parts of speech (POS) and lemma of words; length related features, PMI-IR features, emphasized words and special punctuation symbols. Classification has been performed using support vector machine. Positive-negative emotion classification accuracy is reported to be 60% and that reported for distinct mood labels is 55%.

Emotion analysis on text corpus consisting of 22 children's fairy tales has been performed in (C. O. Alm, D. Roth & R. Sproat., 2005). The classification was performed with three classes, namely, positive emotion, negative emotion, and neutral. Sparse Network of Winnows learning architecture has been used for the classification task with features like first sentence of the story, quote, thematic role type, Wordnet emotion words etc. The F-score reported for neutral, positive emotion and negative emotion classes are 69%, 32% and 13% respectively.

Leshed et al. (G. Leshed & J. J. Kaye., 2006) considered LiveJournal blog posts as the emotion text corpus and perform emotion classification on 50 topmost emotions appearing in the blog posts. The 'bag of word' model of information retrieval combined with tf-idf feature has been used in SVM classifier to assign emotion labels to the blogposts. The average accuracy of the system is reported to be 78%.

Mihalcea et al. (R. Mihalcea & H. Liu., 2006) consider a corpus based approach to classify blog posts from LiveJournal in 'happy' and 'sad' category. Naive Bayes classifier has been used with unigram features for classification task. The accuracy of the system is reported to be 79.13%.

The *a priori* algorithm for association rule mining and Separable Mixture Model (SMM) techniques have been used for text emotion detection in (C. H. Wu, Z. J. Chuang & Y. C. Lin., 2006).   The emotion recognition model was evaluated with a dialog corpus consisting of the students' daily expressions. The model was trained with the dialog corpus and achieved 75.14% precision under a recall rate of 65.67%. The model was then tested with another corpus from different domain (broadcast drama). In this case, the precision was 61.18% at 45.16% recall.

Jung et al. (Y. Jung, H. Park & S. H. Myaeng., 2006) takes a hybrid approach in mood classification in blogposts considering four mood classes: happy, sad, angry and fear. SVM classifier has been used to assign mood labels to documents using the features like term-frequency, n-gram and PMI. The hybrid approach achieves accuracy of 81.80%.

Abasi et al. (A. Abbasi, H. Chen, S. Thoms & T. Fu., 2008) provide an extensive comparison of different features and techniques used for emotion analysis on different corpus and finally propose a support vector regression correlation ensemble (SVRCE) method for text emotion recognition.

The articles taken from periodicals were considered for emotion analysis study in (J. Wang & L Zhang, 2009). Three different models have been constructed in this study: term frequency, semantic characteristics and cognition appraisal theory based models. The micro-average accuracy in cognition appraisal theory based model is reported to be 45%.

3. Limitations of the Previous Works

Based on the study of above mentioned works, following observations can be made.

• Most of the previous works perform emotion analysis in the perspective of the writer, where one text segment portrays only one emotion. On the other hand, one particular text segment can evoke multiple emotions in reader perspective emotion analysis. The multi-labelness of emotion text has not been explored in the previous studies.

• Most of previous studies perform emotion recognition task in document level which may be coarse grained in many applications. Thus finer level analysis like sentence may be explored.

• The performance evaluation measures for multi-label classification are different from that of multi-class or single label classification. So, performance of emotion analyzer should be measured with the metrics defined for multi-label classification.

The above mentioned limitations of the previous works provide the basic motivation behind our work.

4. Multi-Label Emotion Classification Problem and RAKEL

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 $\xi = \{e_i | i = 1, 2, \ldots, |\xi|\}$ be the set of emotion classes (e.g., happy, sad etc.). The task is to find a function $h : S \to 2^\xi$, where 2^ξ is the powerset of ξ.

The problem of reader emotion classification from text data can be mapped to the multi-label text categorization problem. Multi-label classification algorithms have been categorized into two classes: algorithm adaptation methods and problem transformation methods. In algorithm adaptation methods, existing single label classification algorithms are adapted to handle multi-label data whereas, the multi-label data instances are transformed into single label by applying some transformation techniques in problem transformation based methods.

Binary relevance classifier and the Label Powerset classifiers (G. Tsoumakas & I. Katakis., 2007) are examples of problem transformation method. One common problem transformation method is to consider each different subset of ξ as single label. Label Powerset (LP) classifier learns one single classifier $h : S \to 2^\xi$. Random k-Label sets classifier (RAKEL) (G. Tsoumakas & I. Katakis., 2007) builds an ensemble of a number of LP classifiers trained using a different small random subset of ξ. RAKEL has been selected as the representative of problem transformation method in our study as it is reported to outperform the other problem transformation methods and algorithm adaptation methods. Below we provide a brief description of the RAKEL algorithm.

A *k-labelset* is defined as follows:

Definition 1. *A* k-labelset *is defined to be the set* $E \subseteq \xi$ *with* $k = |E|$. *The term* ξ^k *denotes the set of all distinct k-labelsets on* ξ, *where* $|\xi^k| = {}^{|\xi|}C_k$.

The algorithm works in two distinct phases:

• *Ensemble Production:* In this phase, an ensemble of *n* LP classifiers is constructed through *n* iterations. At each iteration, $i = 1, 2, \ldots, n$, a distinct *k*-labelset, E_i, is selected from ξ^k and an LP classifier $h_i : S \to 2^{\xi i}$ is learnt. The parameter *n*, the number of models, is a user specified one and can assume values ranging from 1 to $|\xi^k|$. The permissible range for another user specified parameter, *k*, is 2 to $|\xi| - 1$.

• *Ensemble Combination:* The multi-label classification is performed by combining votes from individual LP classifiers constructed in the first phase. For a test item *t*, each model h_i provides binary decisions $h_i(t, l_j)$ for each label l_j in the corresponding *k*-labelset E_i. Finally the average decision for each label $l_j \in \xi$ is computed. The test instance is labeled with l_j if the average vote is greater than a user specified threshold τ.

5. Emotion Data

The emotion text corpus consists of 1305 sentences extracted from *Times of India* news paper archive. 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 to more than one emotion categories. An example annotation is presented in Table 1.

The statistics related to the gold standard data is provided in term of

• Number of sentences in emotion category

• Label Density (LD): It is defined as the average density of labels and is given by

$$LD = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{|E_i|}{|\xi|} \tag{1}$$

where E_i, the proper subset of ξ, is the label set associated with i^{th} sentence.

• Label Cardinality (LC): It is defined as the average number of labels associated per sentence and is given by

$$LC = \frac{1}{|S|} \sum_{i=1}^{|S|} |E_i| \tag{2}$$

The statistics of the gold standard data set used in this work is presented in Table 2.

6. 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. The extraction of polarity related features involves several steps which are presented below with the following polarity tagged (manual) example sentence (Ne → Negative polarity, P → Positive polarity and no tag implies neutral word).

The [shameful]/Ne [scam]/Ne [stains]/Ne the [clean]/P image of the country.

– STEP 1: In this step, the head words of the verb, subject and object phrases are extracted with the help of the dependency relations obtained by parsing the sentence with Stanford Parser (D. Klein & C. D. Manning., 2003). The output in this step is as follows.

> Subject head word: *scam*
> Verb head word: *stains*
> Object head word: *image*

– STEP 2: The modifier words for verb, subject and object head words are determined by consulting the dependency relations. This step yields the following phrases (M → modifier word, H → head word).

> Subject phrase: *shameful/M scam/H*
> Verb phrase: *stains/H*
> Object phrase: *clean/M image/H*

– STEP 3: The polarity of the verb, subject and object phrases are determined with the help of some rules defined over the modifier word, head word and the dependency relation connecting them. The yield of this step is as follows.

> Subject phrase: *Ne*
> Verb phrase: *Ne*
> Object phrase: *P*

• *Semantic Frame Feature (SF)*: The Berkeley FrameNet project (C. J. Fillmore, 2003) 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 (K. Erk, & S. Padó., 2006). Semantic parsing of an example sentence is presented in Figure 1.

7. Experimental Setup and Results

In this section, we present results of experiments of emotion classification with RAKEL which is an ensemble of a number of base classifiers. The base classifier used in our experiment is C4.5. RAKEL deals with three pre-specified parameters that need to be estimated prior training: i) the number of models, ii) the subset size and iii) the threshold for multi-label output generation. The optimal parameters were estimated via 3-fold cross-validation by varying the number of models from 1 to 500, the subset size from 2 to 3 and the threshold from 0.1 to 0.9 with a step of 0.1. For evaluation 5-fold cross-validation were performed for each experiment.

7.1 Experimental Design

In the multi-label classification framework, we aim at exploring the set of suitable features for emotion recognition. We intend to perform different explorations as follows:

- *Exploration of feature sets:* Experiments on finding proper feature combination out of word, polarity and semantic frame features have been performed.

- *Feature selection:* All the word or semantic frame features are not relevant for emotion classification. Feature selection experiment has been conducted in order to find optimal word and semantic frame feature sets.

7.2 Evaluation Measures

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

- Example based measures (G. Tsoumakas & I. Katakis., 2007): Hamming Loss (HL), Partial match accuracy (P-Acc), Subset accuracy (S-Acc) and F1.

- Ranking based measures (M. L. Zhang & Z. H. Zhou., 2007): One Error (OE), Coverage (COV), Ranking Loss (RL), Average Precision (AVP).

7.3 Comparison of Features

Experiments have been performed with different feature combinations. Table 3 summarizes the results of emotion classification with different features and their combinations with best results presented in bold face.

When the assessment of individual features is concerned, the performance of the emotion classifier with polarity feature (P) deteriorated as compared to the baseline classifier (using word feature (W)) for all the evaluation metrics. This explains how important the terms present in the text are for emotion classification.

On the other hand, use of semantic frames (SF) as features improves the performance of emotion classification significantly. The improvements on partial match accuracy (P-Acc), subset accuracy (S-Acc) and F1 are 6.8%, 9.5% and 5.4% respectively. This significant improvement may be attributed to two different transformations over the word feature set.

- *Dimensionality Reduction:* A significant reduction in the dimension of semantic frame feature set as compared to word feature set has been observed(semantic frame feature dimension = 279 and word feature dimension = 2345).

- *Feature Generalization:* Semantic frame assignment to the terms in the sentences is one generalization technique where conceptually similar terms are grouped into a semantic frame. For example, the terms 'kill', 'assassin' and 'murder' are grouped into a single semantic frame 'Killing'. In semantic frame feature set, unification of these features is performed resulting in less skewedness in feature distribution.

General observations over the feature comparison experiment are as follows.

- The P+SF feature combination performs best in emotion classification with RAKEL. The SF feature performs closer to P+SF as compared to other feature combinations. In case of ranking based measures, the P+SF feature combination outperforms SF by a better margin.

- The polarity feature (P) is inefficient than other combinations but whenever combined with other feature combinations (i.e., W vs. W+P, SF vs. SF+P and W+SF vs. W+SF+P), improvement in   performance has been observed. This improvement can be explained with the fact that the polarity feature may help the word or semantic frame based models by classifying the data set into positive and negative category.

- Whenever W feature is coupled with SF, degradation in performance has been noted (i.e., SF vs. W+SF, P+SF vs. W+P+SF). This degradation in performance is due that fact that SF is a generalization over word feature and introduction of word feature only adds noise to the system.

7.4 Feature Selection

All the words and semantic frame features are not important for emotion classification. So, it is better to filter out the words or semantic frames that are not informative enough for discrimination process. To achieve this we have computed χ^2 statistics (Y. Liu, H. T. Loh & A. Sun., 2009) of the word and semantic frame features. Chi-square measures the lack of independence between a word w and a class e_i and is given by

$$\chi^2(w, e_i) = \frac{|S| [P(w, e_i)P(\overline{w}, \overline{e_i}) - P(w, \overline{e_i})P(\overline{w}, e_i)]^2}{P(w)P(\overline{w})P(e_i)P(\overline{e_i})} \tag{3}$$

The χ^2 value for a word w is given by

$$\chi^2(w) = \sum_{i=1}^{|\xi|} \chi^2(w, e_i) \tag{4}$$

The plot word feature χ^2 value vs. rank (see Figure 2) follows the Zipfian distribution (power law fit with equation $y = \alpha x^{-\beta}$ where $\alpha = 236.43$, $\beta = 0.82$ and goodness of fit $R^2 = 0.89$) having a long tail which is strong indication of feature sparseness problem.

We performed experiment on selecting optimal W feature set size based on their χ^2 values. Top 40% of the total W feature set is found to be optimal feature set. The relative performance after feature selection for W is shown in Figure 3. Similar experiment was performed to select important SF features. 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 Figure 4.

It is evident from results 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 = 0.110, P-Acc = 0.769, F1 = 0.821, S-Acc = 0.670).

7.5 Comparison with Other Systems

The existing methods model the emotion recognition as a single label classification problem. Whereas multi-label classification based approach has been adopted in our approach. As the performance measures for single label classification tasks are different from that of multi-label one, direct comparison of the existing system with ours is not possible. Comparisons may only be performed based on the micro-averaged label based measures like accuracy, precision, recall, and F1 (Tsoumakas, 2007). The comparison of our system with the existing approaches based on these measures is provided in Table 4.

8. Conclusions

In this paper, we have presented a multi-label classification based emotion analysis model. The emotion corpus considered in this study consists of 1305 sentences collected from news archive. An ensemble based multi-label classification technique called *random k-label set (RAKEL)* has been used in our study.

Apart from traditional word feature, we have introduced two other feature groups, namely, polarity based features and semantic frame based features. Experiments with different feature combinations reveal that semantic frame feature combined with polarity based feature performs best in RAKEL framework.

The spurious word or semantic frame features that may not be important in emotion classification task should be removed. To achieve this, we have adopted χ^2 statistics based feature selection strategy. Improvements in performance have been observed after feature selection in both word and semantic frame feature.

References

A. Abbasi, H. Chen, S. Thoms & T. Fu. (2008). Affect analysis of web forums and blogs using correlation ensembles. *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 9, pp. 1168–1180.

A. Valitutti, C. Strapparava & O. Stock. (2004). Developing affective lexical resources. *PsychNology Journal*, vol. 2, no. 1, pp. 61–83, 2004.

B. Pang, & L. Lee. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the ACL 2004,* Barcelona, Spain, Main Volume, pp. 271–278.

C. Fellbaum. (1998). *WordNet An Electronic Lexical Database*. Cambridge, MA ; London: The MIT Press.

C. J. Fillmore (2003) Background to framenet. *International Journal of Lexicography*, vol. 16, no. 3, pp. 235–250.

C. H. Wu, Z. J. Chuang & Y. C. Lin. (2006). Emotion recognition from text using semantic labels and separable mixture models. *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 5, no. 2, pp. 165–183.

C. Ma, H. Prendinger & M. Ishizuka. (2005). Emotion estimation and reasoning based on affective textual interaction. In *ACII* pp. 622–628.

C. O. Alm, D. Roth & R. Sproat. (2005) Emotions from text: machine learning for text-based emotion prediction. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Morristown, NJ, USA: Association for Computational Linguistics, pp. 579–586.

C. Strapparava & R. Mihalcea. (2007). Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on the Semantic Evaluations (SemEval 2007)*, Prague, Czech Republic.

D. Klein & C. D. Manning. (2003). Accurate unlexicalized parsing. In *ACL '03: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, pp. 423–430.

F. R. Chaumartin. (2007). UPAR7: A knowledge-based system for headline sentiment tagging. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*. Prague, Czech Republic: Association for Computational Linguistics, pp. 422–425.

G. Leshed & J. J. Kaye. (2006). Understanding how bloggers feel: recognizing affect in blog posts. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM, pp. 1019–1024.

G. Mishne. (2005). Experiments with mood classification in blog posts. In *Proceedings of the 1st Workshop on Stylistic Analysis of Text for Information Access*.

G. Tsoumakas & I. Katakis. (2007). Multi label classification: An overview. *International Journal of Data Warehouse and Mining*, vol. 3, no. 3, pp. 1–13.

K. Erk, & S. Padó. (2006). Shalmaneser - a toolchain for shallow semantic parsing. In *Proceedings of LREC 2006*, Genoa, Italy, pp. 527–532.

K. H. Y. Lin, C. Yang & H. H. Chen. (2008a). Emotion classification of online news articles from the reader's perspective. In *Web Intelligence*, pp. 220–226.

K. H. Y. Lin, & H. H. Chen. (2008b). Ranking reader emotions using pairwise loss minimization and emotional distribution regression. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. Honolulu, Hawaii: Association for Computational Linguistics, pp. 136–144.

M. L. Zhang & Z. H. Zhou. (2007). ML-kNN: A lazy learning approach to multilabel learning. *Pattern Recognition*, vol. 40, no. 7, pp. 2038-2048.

P. E. Corcoran. (2006). Emotional framing in Australian journalism. In *Australian & New Zealand Communication Association International Conference*. Adelaide, Australia: ANZCA.

P. Katz, M. Singleton & R. Wicentowski. (2007). SWAT-MP: the semeval-2007 systems for task 5 and task 14. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*. Prague, Czech Republic: Association for Computational Linguistics, pp. 308–313.

P. Subasic & A. Huettner. (2001). Affect analysis of text using fuzzy semantic typing. *IEEE Transaction on Fuzzy Systems*, vol. 9, no. 4, pp. 483–496.

R. Mihalcea & H. Liu. (2006). A corpus-based approach to finding happiness. In *AAAI 2006 Symposium on Computational Approaches to Analysing Weblogs*. AAAI Press, pp. 139–144.

Y. H. Cho & K. J. Lee. (2006). Automatic affect recognition using natural language processing techniques and manually built affect lexicon. *IEICE - Transactions on Information and Systems*, vol. E89-D, no. 12, pp. 2964–2971.

Y. Jung, H. Park & S. H. Myaeng. (2006). A hybrid mood classification approach for blog text. In *PRICAI*, pp. 1099–1103.

Y. Liu, H. T. Loh & A. Sun (2009). Imbalanced text classification: A term weighting approach. *Expert systems with Applications,* vol. 36, no. 1, pp. 690-701.

Z. Kovecses. (2003). Language and emotion concepts. In *Metaphor and Emotion: Language, Culture, and Body in Human Feeling*. Cambridge: Cambridge University Press.

Z. Kozareva, B. Navarro, S. Vazquez & A. Montoyo. (2007). UA-ZBSA: A headline emotion classification through web information. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*. Prague, Czech Republic: Association for Computational Linguistics, pp. 334–337.

Table 1. An example annotation (1 → Emotion evoked in reader, 0 → Emotion not evoked)

Sentence	Disgust	Fear	Happiness	Sadness
The four terrorists in the Taj Mahal hotel have killed virtually anyone and everyone they saw.	0	1	0	1

Table 2. Statistics for emotion data

Sentence Distribution		LC	LD
Class	No of Sentences		
Disgust	307		
Fear	371		
Happiness	282	1.30	0.26
Sadness	735		

Table 3. Comparison of features (W → Word feature, P → Polarity feature and SF → Semantic frame feature)

Measure	W	P	SF	W+P	P+SF	W+SF	W+P+SF
Example based measures							
HL	0.166	0.237	**0.112**	0.155	**0.112**	0.123	0.129
P-Acc	0.685	0.531	0.753	0.710	**0.764**	0.752	0.748
F1	0.744	0.589	0.798	0.766	**0.812**	0.809	0.806
S-Acc	0.566	0.414	0.661	0.597	**0.666**	0.635	0.627
Ranking based measures							
OE	0.343	0.342	0.156	0.215	**0.146**	0.171	0.169
COV	0.910	1.119	0.672	0.785	**0.663**	0.684	0.684
RL	0.235	0.347	0.203	0.241	**0.194**	0.198	0.199
AVP	0.796	0.770	0.893	0.857	**0.898**	0.887	0.887

Table 4. Comparison of proposed system with other emotion classification systems

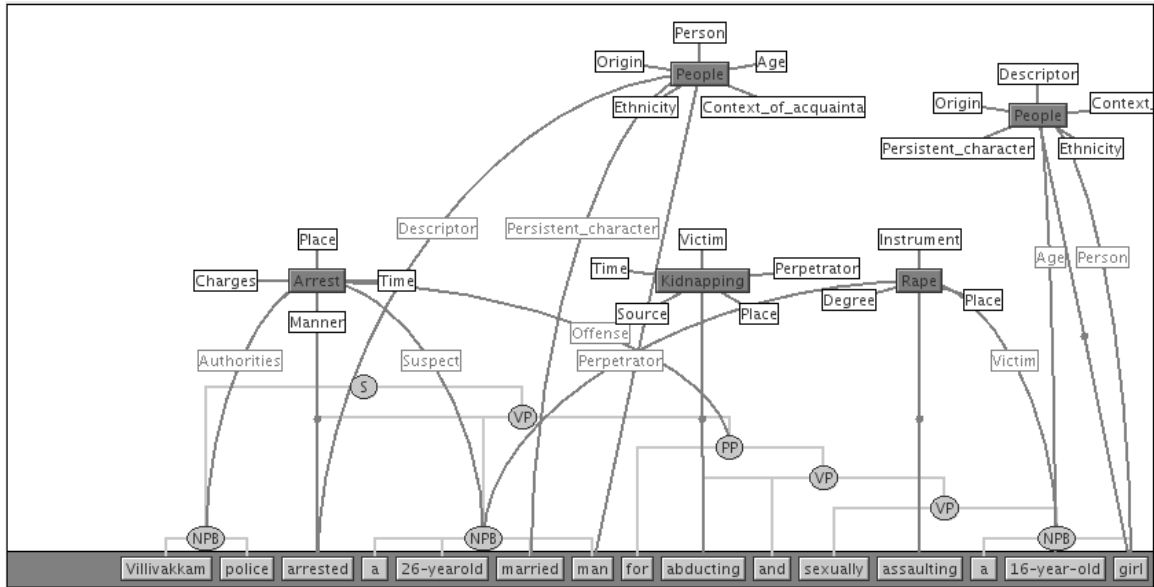| System	Accuracy	Precision	Recall	F1
UPAR7	89.43	27.56	5.69	9.43
UA-ZBSA	85.72	17.83	11.27	13.81
SWAT	88.58	19.46	8.62	11.95
Li and Chen	76.88	--	--	--
Our System	**88.2**	**84.42**	**79.93**	**82.1**

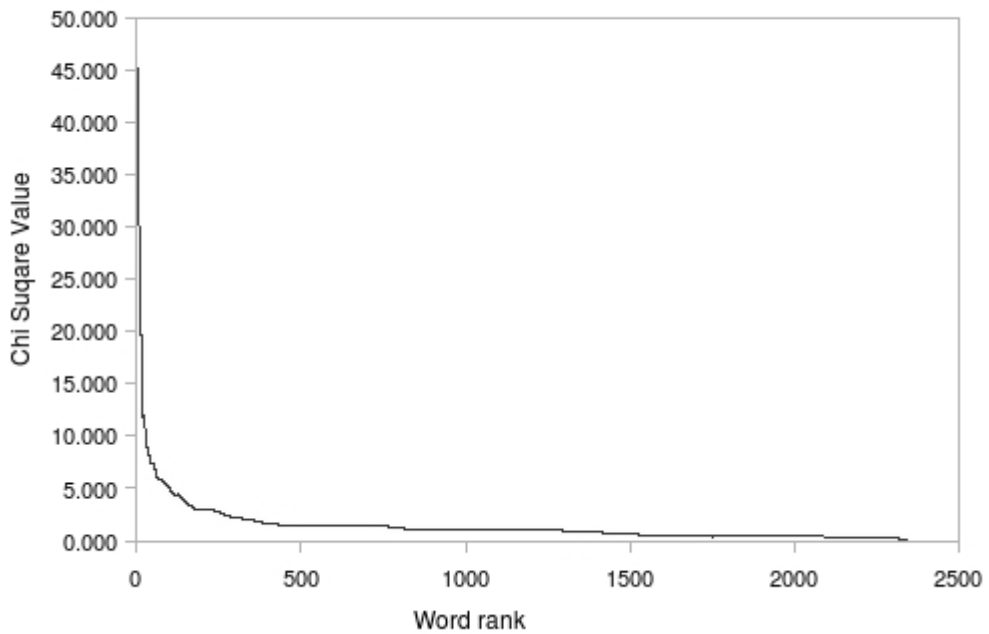Figure 1. Semantic parsing of an example sentence

Figure 2. χ^2 vs. rank plot for word feature

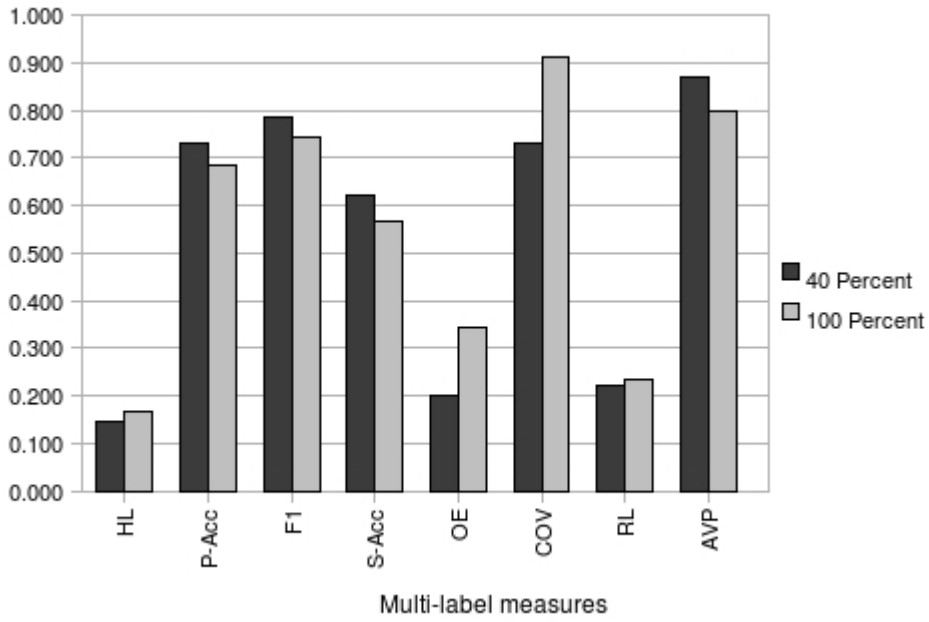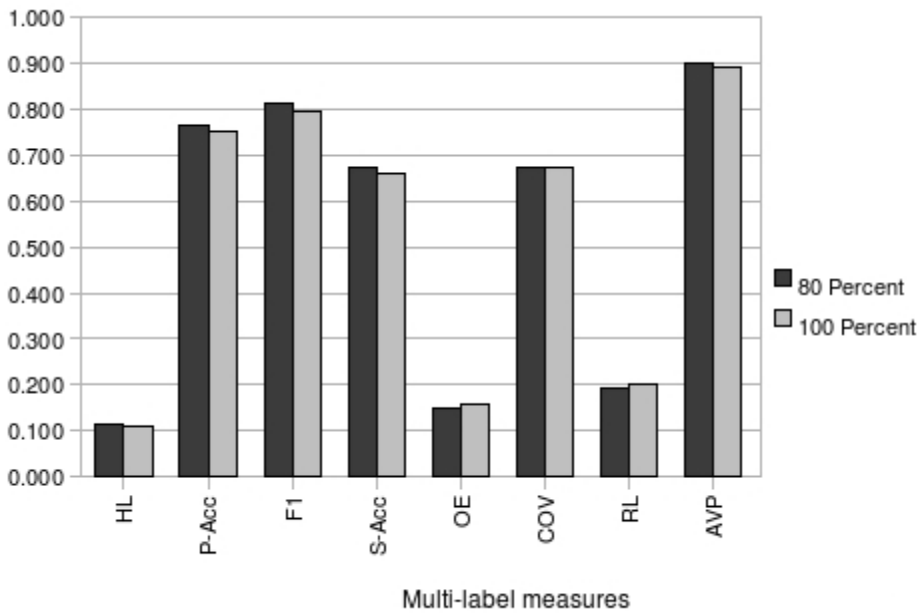Figure 3. Relative performance after χ^2 word feature selection

Figure 4. Relative performance after χ^2 semantic frame feature selection