

Identifying Emotional Expressions, Intensities and Sentence level Emotion Tags using a Supervised Framework*

Dipankar Das and Sivaji Bandyopadhyay

Department of Computer Science and Engineering, Jadavpur University,
188, Raja S.C. Mullick Road, Kolkata -700032, West Bengal
dipankar.dipnil2005@gmail.com, sivaji_cse_ju@yahoo.com

Abstract. The present work deals with the extraction of emotional expressions and tagging of English blog sentences with Ekman's six basic emotion tags and any of the three intensities: low, medium and high. Baseline system is developed based on *WordNet Affect* lists and dependency relations. Support Vector Machine (SVM) based supervised framework is employed by incorporating different word and context level features. The feature analysis is carried out on 358 development sentences followed by Information Gain Based Pruning. Application of admissible tag sequences and a class-splitting technique improves the system's performance and reduces the label bias problem of SVM. The supervised system outperforms the baseline system and achieves average *F-Scores* of 82.72%, 76.74% and 89.21% for emotional expressions, sentential emotion tags and intensities respectively on 565 gold standard test sentences. A comparative evaluation shows that sentential emotion tagging based on emotional expressions, intensities and context features bridges the gap of identifying sentential emotion depending only on words.

Keywords: Emotional Expression, Intensity, Negation, Conjunctions, Information Gain.

1 Introduction

In psychology and common use, emotion is an aspect of a person's mental state of being, normally based in or tied to the person's internal (physical) and external (social) sensory feeling (Zhang *et al.*, 2008). Emotion identification from natural language texts has drawn the attention of several information processing communities, e.g. reviews (Turney, 2002), news (Lin *et al.*, 2007), Question Answering (Wilson *et al.*, 2005), Information Retrieval (Pang and Lee, 2008).

Several efforts have been made by the natural language processing researchers to identify emotion at different level of granularities such as word, sentence or document (Ku *et al.*, 2006; Das and Bandyopadhyay, 2009b; Das and Bandyopadhyay, 2010). But in many domains of text, the values of individual phrases may bear little relation to the overall sentiment expressed by the text. Just like words, phrases are considered as the informative and emotion expressive units of any sentence and are used in identifying document level emotion tags (McDonald *et al.*, 2007). Hence, the current dependence on individual word level information to identify sentential emotion has motivated the search for phrase level clues.

The present task extracts the emotional expressions (word/phrase) from each of the English blog sentences and assigns any of the Ekman's (1993) six basic emotion tags along with any of the three intensity types: low, medium and high to each sentence. Blog is one of the crucial,

* The work reported in this paper was supported by a grant from the India-Japan Cooperative Programme (DST-JST) 2009 Research project entitled "Sentiment Analysis where AI meets Psychology" funded by Department of Science and Technology (DST), Government of India.

communicative and informative repository of text based emotional contents in the Web 2.0 (Lin *et al.*, 2007). The English blog corpus (Aman and Szpakowicz, 2007) annotated with emotional expressions, sentential emotion tags and intensities is considered to accomplish our present task.

The baseline system identifies emotional expressions, sentential emotion tags and intensities. Sentences are passed through the Stanford Dependency Parser (Marneffe *et al.*, 2006) to identify the dependency relations. Emotional expressions are identified using *WordNet Affect* (Strapparava and Valitutti, 2004). Different types of dependency relations of the parsed sentences provide positional, as well as intensity related hints to the emotional expressions. The evaluation on 565 gold standard test sentences gives the *F-Scores* of 62.44%, 56.97% and 79.01% for emotional expressions, sentential emotion tags and intensities respectively.

The best emotion feature set for the Support Vector Machine (SVM) based supervised framework is identified based on the *F-Scores* obtained from 358 development sentences. The features like emotion words, intensifiers, negations, conjuncts and discourse markers obtain high Information Gain (*InfoGain*) on the development sentences. The unnecessary words, which carry no emotion features, yield low Information Gain (*InfoGain*) and are filtered out. The performance of the supervised system is further improved by identifying the admissible tag sequences and a class splitting technique. Emotion pairs like *happy-surprise*, *angry-disgust*, and *angry-sad* and intensity pairs like *high-medium* and *medium-low* show the problem in sentential emotion and intensity tagging. Adding a special feature of emotional composition that identifies the closeness between two emotion types solves the problem. Overall, the supervised system achieves average *F-Scores* of 82.72%, 76.74% and 89.21% for emotional expressions, sentential emotion tags and intensities on 565 gold standard test sentences respectively. The supervised system outperforms the baseline system significantly and also performs better in terms of *F-scores* compared to the accuracy values of the word level system (Das and Bandyopadhyay, 2009a).

The rest of the paper is organized as follows. Section 2 surveys the related work. Pre-processing is described in Section 3. The baseline system is discussed in Section 4. Section 5 describes a supervised framework for identifying the emotional components. Experimental results are discussed in Section 6. Finally Section 7 concludes the paper.

2 Related Work

The earlier works on polarity shifters and prior or contextual polarity approaches for sentiment analysis are described in (Polanyi and Zaenen, 2004; Kim and Hovy, 2004). Another related work described in (Wilson *et al.*, 2005) gives the ground to analyze sentiment at phrase levels. The tasks described in (Meena and Prabhakar, 2007) determine the sentential sentiment based on the phrase level information considering the impact of conjuncts or intensifiers. Majority of the above studies are carried out for sentiment expressions whereas our present approach aims to identify emotional expressions and assigns the emotion tags and intensities to the sentences.

In order to estimate affects in text, the model proposed in (Neviarouskaya *et al.*, 2007) processes symbolic cues and employs natural language processing techniques for word/phrase/sentence level analysis, considering relations among words in a sentence. The current trend in the emotion analysis area is exploring machine learning techniques (Sebastiani, 2002), which considers the problem as text categorization or analogous to topic classification and underscores the difference between machine learning methods and human-produced baseline models (Alm *et al.*, 2005). As large blog data set is suitable for machine learning models, several supervised and unsupervised machine learning classification techniques on blog data for comparative evaluation have been carried out (Mishne and Rijke 2006). Affective text shared task on news headlines at SemEval 2007 for emotion and valence level identification (Strapparava and Mihalcea, 2007) has drawn the focus to this field. In contrast, our present task extracts the emotional expressions and identifies the sentential emotion tags and intensities based

on intensifier, negations and conjuncts. The incorporation of Information Gain Based pruning (IGBP) and class splitting techniques are the contributions in present task.

3 Pre-Processing

In the present study, we have used the English blog corpus described in (Aman and Szpakowicz, 2007). The sentences are tagged with single or multiple emotion tags and intensities. Emotional expressions are also annotated in the corpus. The whole corpus is divided into training, development and test sets containing 2500, 300 and 500 sentences with single emotion tags respectively. In addition to that, 200, 58 and 65 sentences of mixed emotion types are considered for training, development and test respectively.

The emotion word lists, *WordNet Affect* (Strapparava and Valitutti, 2004) is available for Ekman's six basic emotions in English. The six lists are used to obtain the emotion or affect words present in the emotional expressions that in turn contribute for identifying sentential emotion tags. The blog corpus contains different emoticons (☹, ☺, ☺) that represent different types of emotions. No prior knowledge was given for the emoticons in the corpus. Hence, the emoticons are replaced by the appropriate emotion words using a manually prepared knowledgebase.

The input sentences are passed through the Stanford Parser (Marneffe *et al.*, 2006), a probabilistic lexicalized parser containing 45 different part of speech (POS) tags from the Penn Treebank tagset. Different types of dependency relations and POS tags are retrieved successively from the parsed data. The parsed results are used in the baseline and supervised systems.

4 Baseline System

The overall emotion expressed in a sentence depends on the emotional constituents present in the phrases or in the words along with other surrounding contextual hints (Neviarouskaya *et al.*, 2007). The emotional components present in an emotional expression are represented as an anchoring vector. Each slot of the vector is filled by the information extracted from the emotional expression such as emotion or affect word, intensifier, negation and conjuncts.

4.1 Extracting Emotional Expressions

We extract all component words from the dependency relations that contain at least one emotion word, and match them against the *WordNet Affect* lists. The words present in the extracted relations are then treated as candidate seeds for the anchoring vectors representing emotional expressions. Identification of an emotional expression containing a single emotion word is straightforward. But, we include the positional information about words from the dependency relations in order to identify long emotional expressions. Consecutive words that appear in the dependency relations and contain at least one emotion word are identified as an emotional expression. For example, in the following sentence, the dependency relations containing the emotion word "angry" and the consecutive component words "really", "angry" and "rant" with position number 9, 10 and 11 constitute the underlined emotional phrase.

"But the rest of it came across as really angry rant". [advmod(angry-10, really-9), amod(rant-11, angry-10)]

In many cases, the components of a given emotional expression are separated by articles (e.g. a, an, the) conjuncts (e.g. and, or, but etc.), negations (e.g. not, neither etc.) or intensifiers (so, very, less, much). The aim is to accumulate the words for constructing the emotional expressions based on the positional clues with maximum coverage of the contributory components. Based on empirical testing, the limit for the maximum difference between the successive positions is set at 2~3. The technique is used for identifying longer emotional expressions. For example, in the following sentence, two emotional words "angry" and "frustrated" are not consecutive but are joined by a conjunct "or" and separated by 2 positions.

“We feel ***angry or frustrated*** with others or ourselves”. [acomp(feel-2, *angry-3*),
acomp(feel-2, *frustrated-5*), conj_*or* (*angry-3, frustrated-5*)]

In addition to the dependency relations containing emotion words alone, other dependency relations such as advmod()[adverbial modifier], amod() [adjectival modifier], neg [negation modifier], conj_() [conjunct] are also extracted as these relations carry emotional intensifiers that contribute towards tagging the sentences with emotion tags and intensity. Each emotional expression is tagged with the *WordNet Affect* classes in which the words of the emotional expression occur. Each of the sentences is also tagged with Ekman’s six emotion tags that are associated with its corresponding emotional expressions. Each sentence may contain more than one emotion tag. Otherwise, the sentences are treated as neutral sentences.

4.2 Identification of Intensity

Three types of intensities (I) (e.g. low (L), medium (M) and high (H)) have been considered in the present task. Two types of intensifiers (INTF) such as positive and negative are considered for assigning intensities to the sentences. The intensifiers may be present inside an emotional expression. The POS categories of the preceding and following elements of the emotion word in an emotional expression are identified from the parsed sentences. The JJ (adjective) and RB (adverb) tagged elements are considered as the candidate of intensifiers. To identify the polarity of an intensifier, each of them is searched in the *SentiWordNet* (Esuli and Sebastiani, 2006). If the candidate word is found in the *SentiWordNet*, then the positive and negative scores of the word are retrieved from the *SentiWordNet*. The intensifier is classified as either positive (pos) or negative (neg) for which the average retrieved score is higher. A list of commonly used negative words has been prepared manually. The words present in the *neg* [negation modifier] type dependency relation are also considered as the negative words. Some rules (as shown in Table 1) are applied to capture the role of different intensifiers and negations in marking the sentences with post emotion tags along with intensity.

Table 1: Rules for tagging an emotional anchoring vector with intensity.

(Sentence with prior tag [EW])[Example]	post tag [EW_I]
([EW])[happy]	[EW_M]
(INTFpos, [EW]) [very, happy]	[EW_H]
(INTFneg, [EW]) [less, surprised]	[EW_L]
(NEG, [EW]) [not, happy]	[EW_L]
(NEG, INTFpos, [EW]) [not, very, happy]	[EW_M]
(NEG, INTFneg, [EW]) [not, less, disappointed]	[EW_H]

4.3 Sentence Level Tagging

Each sentence is assigned with the emotion tags according to its corresponding emotional expressions. The emotion tags are considered as prior tags to each sentence. The final sentential emotion tagging is done based on the intensity assignment along with the role of conjuncts.

Rule based strategies have been adopted to cope with conjuncts. Emotional expressions containing conjuncts do not change the sentential emotion tags given for the corresponding emotional expressions. In the following example, the sequence of negation (nothing), conjunct (but) and emotion word (sorry) does not change the emotion tag, “sad”. Robert feels ***nothing but sorry*** for the common people”. Sometimes conjuncts present between two or more emotional expressions do not play any role to change the sentential emotion tags that were already associated with the sentence. For example, “They were ***very angry*** at that moment ***but*** you managed the situation with your ***impressive quality***.” The rules are shown in Table 2. The symbol “X” stands for any of the three intensities as identified for the emotional expressions.

The baseline system achieves the average *F-Scores* of 62.44%, 56.97% and 79.01% for emotional expressions, sentential emotion tags and intensities respectively on 500 gold standard test sentences of single emotion tags and 65 sentences of multiple emotion tags. The detailed *precision, recall* and *F-Scores* are shown in Table 5.

Table 2: Rules for conjuncts for sentence level emotion tagging.

Rules for Tagging (Tag EW _{i≠j})[Example]	Final Tag(s)(Tag EW_I)
NEG, CONJ, (EW)[nothing, but, sorry]	(EW_X)
(EW _i), CONJ, (EW _j)	(EW _i _X, EW _j _X)

5 Supervised Framework

The Support Vector Machine (SVM) (Cortes and Vapnik, 1995) based supervised framework has been used to extract the emotional expressions as well as to tag the sentences with emotions and intensities. The emotion and intensity tagging is done at sentence level. For emotional expressions, the task is to label any of the six emotion tags to a single word or a sequence of words in a sentence. Other words are tagged as neutral (*emo_ntrl*). The sentential emotion and intensity tagging to the sentences is based on the emotional expressions along with other discourse level features. First a set of standard pre-processing steps is carried out, *viz.*, *tokenizing, stemming* and *stop word removal*. Tools provided by *Rapidminer's*¹ text plug in were used for these tasks. Stemming is done by WordNet's morphological analyzer. The system is trained with 2700 sentences. The best feature set is identified using the 358 development sentences. Different unigram and bi-gram context features (word, POS tag, Intensifier, negation) and their combinations were generated from the training corpus. The Information Gain Based Pruning (IGBP), Admissible Tag Sequence (ATS), Class Splitting technique and Emotional Composition features are applied on the development set and it improves the performance of the supervised system significantly. The detail results are shown in Table 4. All the results are obtained by 10 fold cross validation method.

5.1 Features for Training

Feature plays a crucial rule in the SVM framework. By manually reviewing the blog data and different language specific characteristics, 11 active features, word level as well as context level, have been selected heuristically for our classification task. Each feature with discrete value is adopted for intensity at the word level.

- **Emotion/Affect Words (EW):** The presence of a word in the *WordNet Affect* lists (Strapparava and Valitutti, 2004) identifies the emotion/affect words.
- **Parts-of-Speech (POS):** We are interested with the *verb, noun, adjective* and *adverb* words as these are emotion informative constituents.
- **Intensifiers (INTF):** Dependency relations such as *amod()* [adjectival modifier], *advmod()*[adverbial modifier] containing JJ (adjective) and RB (adverb) tagged elements are considered as intensifiers (“and he said he just doesn't have *enough* time to love”). If the intensifier is found in the *SentiWordNet*, then the positive and negative scores of the intensifier are retrieved from the *SentiWordNet*. The intensifier is classified into the list of positive (pos) (**INTFpos**) or negative (neg) (**INTFneg**) for which the average retrieved score is higher (*very, less, few, enough, little, much* etc.).
- **Direct and Transitive Dependency relations (DD and TD):** The direct dependency (DD) is identified based on the simultaneous presence of the emotion word and the other word in the same dependency relation whereas the transitive dependencies (TD) are verified if they are connected via one or more intermediate dependency relations.

¹ <http://rapid-i.com/content/blogcategory/38/69/>

- **Negations (NEG):** Dependency relations such as *neg_()* [negation modifier] or the negative words from a manually prepared list (*no, not, nor, neither* etc.), e.g., “there's *no* way I'm turning it down”, are adopted for *Negations* feature..
- **Conjuncts (CONJ):** The dependency relation *conj_()* [conjunct] identifies the *Conjuncts* features (*and, or, but*), e.g., “*But* I asked him about WHY he doesn't cook for himself”.
- **Punctuation Symbols (Sym):** Symbols such as comma (,), (!), (?) are often used in single or multiple numbers to emphasize emotional expressions and considered as crucial clues for identifying emotional presence (“I can't believe she is FINALLY here!!!”).
- **Discourse Markers (DM):** The present task aims to identify only the explicit *discourse markers* that are tagged by *conjunctive_()* or *mark_()* type dependency relations of the parsed constituents (e.g. *as, because, while, whereas*).
- **Causal Verbs (CV):** The lists for *causal* verbs are prepared by processing the XML files of English VerbNet (Kipper-Schuler, 2005). If a class contains any frame with semantic type as *Cause*, we collect the member verbs from that XML class file. The list contains a total of 250 *causal* verbs (e.g. *cause, happen, occur* etc.).
- **Capitalized Phrases (CP):** A capitalized word or a long capitalized phrase segment, e.g., “I forgot how demeaning *BME* classes are”, “Terrorists *MAKE ME SICK*, they ought to all be horrifically detained” is considered as the *Capitalized Phrases* feature.
- **Emoticons (emot_icon):** The emoticons (☺, ☹, 😊) and their consecutive occurrence generally contribute as much as real sentiment to the words that precede or follow it.

5.2 Feature Selection

The combination of multiple features in comparison with a single feature generally shows a reasonable performance enhancement of any classification system. The impact of different features and their combinations were measured on the development set and the results are given in Table 3. We added each feature into the active feature list one at a time if the inclusion of the feature in the existing feature set improves the *F-Score* of the system on the development set. The final active feature set has been applied on the test data.

The Emotion Word, POS, Intensifier and direct dependency features played an important role in extracting emotional expressions as well as tagging sentences with emotions and intensities. During the testing on development set, certain features (e.g. transitive dependency relations, causal verbs, discourse markers) play an effective role in sentential emotion tagging but degrade the performance of extracting emotional expressions. During SVM-based training phase, the current token word with three previous and three next words and their corresponding POS along with negation or intensifier were selected as context feature for that word. Evaluation results on the development set of 358 sentences demonstrated *F-Score* values of 76.07%, 64.23% and 81.04% for emotional expressions, sentential emotion tags and intensities respectively.

5.3 Information Gain Based Pruning (IGBP)

The importance of incorporating the features is examined through Information Gain (*InfoGain*). This decision technique is used to measure the importance of a feature (X) with respect to the class attribute (Y). Formally, information gain of a feature X with respect to a class attribute Y is the reduction in uncertainty about the value of Y when we know the value of X.

$$InfoGain(Y;X) = entropy(Y) - entropy(Y|X)$$

where X and Y are discrete variables taking values $\{x_1, x_2, \dots, x_m\}$ and $\{y_1, y_2, \dots, y_n\}$ respectively. The Entropy (Y) is defined as

$$Entropy(Y) = - \sum_{i=1}^n P(Y = y_i) \log_2 P(Y = y_i)$$

The conditional entropy of Y given X is defined as

$$Entropy(Y|X) = - \sum_{j=1}^m P(X = x_j) Entropy(Y|X = x_j)$$

Features with high Information Gain reduce the uncertainty about the class to the maximum. In our experiment on the development set, all the features except causal verbs and transitive dependency relations achieve an Information Gain above 50%. Hence, the information Gain of (50%) has been fixed as Information Gain (*InfoGain*) threshold. The word features (e.g. non-emotional words such as *gather*, *seem*) are filtered from the corpus based on the Information Gain (*InfoGain*) threshold.

Table 3: *F-Scores* (in %) of some crucial feature(s) applied on the development set.

Feature(s)	Emotional Expression	Sentential Emotion	Intensity
EW	42.45	46.65	-
EW + POS	44.22	48.43	-
INTF _{pos} + INTF _{neg}	-	-	56.78
INTF _{pos} +neg+NEG	-	-	60.09
EW+POS+INTF	64.78	50.03	66.05
EW+POS+INTF+DD	67.35	51.82	68.32
EW+POS+INTF+TD	66.97	52.24	68.03
EW+POS+INTF+DD+NEG	70.67	52.58	72.66
EW+POS+INTF+DD+NEG+CONJ	72.08	54.77	74.54
EW+POS+INTF+DD+NEG+CONJ+Sym	74.87	56.34	76.29
EW+POS+INTF+DD+NEG+CONJ+Sym+DM	74.92	58.02	78.11
EW+POS+INTF+DD+NEG+CONJ+Sym+DM+CV	73.44	61.21	79.76
EW+POS+INTF+DD+NEG+CONJ+Sym+DM+CP+emot_icons	74.67	62.33	80.04
EW+POS+INTF+DD+NEG+CONJ+Sym+DM+CP+Contex Features	76.07	64.23	81.04

5.4 Assigning Admissible Tag Sequence (ATS)

During evaluation on the development sentences, it is found that the classifier produces a sequence of inadmissible classes (e.g., *Happy* immediately followed by *Sad* or *high* followed by *low*) to a consecutive word sequence. To eliminate such sequences, we define a transition probability between word classes $P(c_i | c_j)$ to be equal to 1 if the sequence is admissible, and 0 otherwise. The probability of the classes c_1, \dots, c_n is assigned to the words in a sentence “ s ” in an emotion class D is defined in the following equation

$$P(c_1, \dots, c_n | s, D) = \prod_{i=1}^n P(c_i | s, D) * P(c_i | c_{i-1})$$

where $(c_i | s, D)$ is determined by the SVM classifier.

5.5 Applying Class Splitting Technique

A close investigation for the emotional expressions on development set suggests that the errors are mostly due to the uneven distribution between emotion and non-emotion tags. The results on the development set are shown using confusion matrix in Table 7. The number of non-emotional or neutral type tags is comparatively higher than that of emotional tags of emotional expressions.

So, one solution to this unbalanced class distribution is to split the ‘non-emotion’ (emo_ntrl) class into several subclasses effectively. That is, given a POS tagset POS , we generate new emotion classes, ‘emo_ntrl-C’|C \in POS . We have forty-five (45) subclasses in the English POS tagset, which correspond, to non-emotion regions such as ‘emo_ntrl-NN’ (common noun), ‘emo_ntrl-VFM’ (verb finite main), ‘emo_ntrl-JJ’ (adjective) etc. Evaluation results of the system have shown the improvement for extracting emotional expressions sentential emotion tagging.

5.6 Adding Emotional Composition Feature

The errors for sentential emotion tagging and intensity tagging are mostly due to the uneven distribution between some closely related emotion classes such as *happy-surprise*, *angry-disgust*, and *angry-sad* and intensity classes such as *high-medium* and *medium-low*. The confusion matrix for emotion and intensity classes are shown in Table 6 and Table 7 respectively. It is observed that the co-occurrence of this tag pairs is comparatively higher than other tag pairs. Hence, we have added a special feature for emotional composition that identifies the closeness of emotion classes. For example, if any of the two tags *happy* and *surprise* is tagged in a sentence, the emotional composition feature is to be flagged for the other. This feature improves the performance of sentential emotion and intensity tagging.

Table 4: Improvement *F-Scores* (in %) of the supervised system on development set.

	Emotional Expression	Sentential Emotion	Intensity
Single Emotion (500)	75.80	64.09	80.99
Mixed Emotion (65)	76.04	65.37	79.21
After IGBP	77.01	66.74	82.43
After ATS	79.33	69.14	83.75
After Class Splitting	81.44	73.55	86.08
Emotional Composition	82.72	76.74	89.21

6 Experimental Results

Each of the 500 gold standard test sentences contains an emotional expression, a single emotion tag and intensity. Other 65 gold test sentences contain multiple sentential emotion tags, emotional expressions and intensities. The sentential emotion tagging is done based on the majority voting of the emotion tags identified for emotional expressions. In case of a tie in majority voting, the intensities associated with the sentences are considered for assigning the final tag. The preferred ordering sequence is *high* > *medium* > *low* which means that the emotion tag with *high* intensity is assigned at the sentence level over the emotion tag with *medium* intensity and so on. A noticeable improvement of precision is achieved for sentential emotion tagging. It is observed that the sentences with *high* intensity type improve the *precision* rather than *low* or *medium* type. The reason may be more explicit usage of *high* intensity than the others. The individual contributions of three types of intensities in calculating the precision score are shown in Table 5. It is found that the intensities high (H) and general (G) types contribute reasonably than low (L) category. It signifies the fact that the system satisfactorily identifies the sentential emotion types from its associated intensity clues. In case of sentences containing multiple emotion types, all the emotion tags are considered as the probable candidates for assigning sentential emotion tags.

The *recall* (R) value for sentential emotion tagging is measured by counting the number of sentences that are assigned with at least one sentential emotion tag. The *precision* (P) value is calculated by counting the number of sentences for which the assigned sentential emotion tags match with their corresponding gold standard annotated emotion tags. *F-Score* is defined as $(2 \cdot P \cdot R) / (P + R)$. The similar measures of *precision*, *recall* and *F-Score* are considered for evaluating intensity tagging and the results with respect to all emotion classes for the baseline and supervised systems are shown in Table 5. In comparison with *recall*, the loss in *precision* has occurred due to the frequent use of metaphoric words in blogs as the metaphors are hard to tag with their emotional senses. The supervised system performs significantly better than the baseline system as the baseline system suffers for the sentences that either start with conjuncts (CONJ) or the sentences where the negations (NEG) are present far apart from the identified emotional expressions (e.g. *But, the fully sad story can give no enough detail*”).

The task described in (Das and Bandyopadhyay, 2009a) deals only with words whereas our present task integrates the knowledge of context features (e.g. intensifier, negation, conjuncts) into account. The comparison of the *F-Score* values on sentential emotion tags based on emotional expressions with the accuracies obtained on sentential emotion tagging based on word level (Das and Bandyopadhyay, 2009a) is shown in Table 5. It shows that the present system produces significantly better results. The reason may be the integration of context level features such as intensity, negation, conjuncts, discourse markers etc.

Table 5: Precision (P), Recall (R) and F-Scores (F) (in %) of baseline and supervised systems on test set

	Emotional Expression			Sentential Emotion			Intensity		
	P	R	F	P	R	F	P (H, M, L)	R	F
Baseline Single Emotion (500)	.60	.64	.62	.53	.60	.56	.77 (.74, .03, .00)	.81	.79
Baseline Mixed Emotion (65)	.57	.65	.61	.51	.53	.52	.72 (.68, .03, .01)	.77	.74
Supervised Single Emotion (500)	.80	.85	.82	.71	.82	.76	.82 (.77, .03, .02)	.95	.89
Supervised Mixed Emotion (65)	.73	.85	.79	.72	.78	.75	.80 (.75, .02, .03)	.94	.87
Word Based				.6046 (Acc.)					

Table 6: Confusion Matrix for word level emotion tagging distribution (in %) on development set.

	anger	disgust	fear	happy	sad	surprise	emo_ntrl
anger	0.64	0.07	0.02	0.00	0.06	0.00	0.22
disgust	0.08	0.65	0.00	0.00	0.02	0.00	0.25
fear	0.02	0.01	0.61	0.00	0.05	0.00	0.28
happy	0.00	0.00	0.00	0.68	0.00	0.07	0.25
sad	0.07	0.00	0.05	0.00	0.67	0.00	0.21
surprise	0.00	0.00	0.01	0.12	0.00	0.66	0.21
emo_ntrl	0.00	0.00	0.00	0.00	0.00	0.00	0.87

Table 7: Confusion Matrix for Intensity Tagging distribution (in %) on development set.

	High (H)	Medium (M)	Low (L)
High (H)	0.75	0.24	0.01
Medium (M)	0.02	0.85	0.13
Low (L)	0.00	0.21	0.79

7 Conclusion

The automatic extraction of emotional expressions, sentential emotion tags and intensities from English blog data is done in the present task. From the overall analysis, it is observed that the identification of intensities is helpful in tagging sentential emotions. The handling of metaphors and their impact in detecting sentence level emotion is not considered. Future analysis concerning the sentence level emotion can be used for document level emotion identification.

References

- Alm Cecilia Ovesdotter, Dan Roth, Richard Sproat. 2005. Emotions from text: machine learning for text-based emotion prediction. *HLT- EMNLP*, pp. 579 - 586, Canada.
- Cortes C. and V. Vapnik. 1995. Support-Vector Network. *Machine Learning*, 20, pp. 273–297

- Das. D and S. Bandyopadhyay. 2009a. Emotion Tagging – A Comparative Study on Bengali and English Blogs . *ICON-09*, pp. 177-184, India.
- Das. D and S. Bandyopadhyay. 2009b. Word to Sentence Level Emotion Tagging for Bengali Blogs. *ACL-IJCNLP 2009*, pp. 149-152, Singapore.
- Das. D and S. Bandyopadhyay. 2010. Sentence Level Emotion Tagging on Blog and News Corpora. *Journal of Intelligent System (JIS)*, vol. 19(2), pp. 125-134, 2010.
- Ekman Paul. 1993. Facial expression and emotion. *American Psychologist*, 48(4):384–392.
- Esuli Andrea and Fabrizio Sebastiani. 2006. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. *LREC-06*.
- Kipper-Schuler K. 2005. VerbNet: A broad-coverage, comprehensive verb lexicon, Ph.D. thesis, Computer and Information Science Dept., University of Pennsylvania, Philadelphia, PA.
- Kim S-M. and E. Hovy. 2004. Determining the sentiment of opinions. In *Coling 2004*.
- Ku Lun-Wei, Yu-Ting Liang, and Hsin-Hsi Chen. 2006. Opinion extraction, summarization and tracking in news and blog corpora. *AAAI-2006*, pp. 100-107.
- Lin K. H.-Y., Yang C. and Chen H.-H. 2007. What emotions news articles trigger in their readers? *SIGIR*, 733-734.
- Marneffe Marie-Catherine de, Bill MacCartney, and Christopher D. Manning. 2006. Generating Typed Dependency Parses from Phrase Structure Parses. *5th LRE*.
- McDonald Ryan, Kerry Hannan Tyler Neylon Mike Wells Jeff Reynar. 2007. Structured Models for Fine-to-Coarse Sentiment Analysis. *ACL*.
- Meena Arun and T.V. Prabhakar. 2007. Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis. *ECIR 2007*, LNCS 4425, pp. 573 – 580.
- Mishne G. and M. de Rijke. 2006. Capturing Global Mood Levels using Blog Posts. *Proceedings of AAI, Symposium on Computational Approaches to Analysing Weblogs*, pp. 145-152.
- Neviarouskaya, A., Prendinger, H., and Ishizuka, M. 2007. Narrowing the Social Gap among People Involved in Global Dialog: Automatic Emotion Detection in Blog Posts, *ICWSM'2007*.
- Pang Bo, and Lillian Lee. 2008. Foundations and Trends in Information Retrieval 2, vol. 1-2, pp. 1–135.
- Polanyi L. and A. Zaenen. 2004. Contextual valence shifters. Computing Attitude and Affect in Text: Theory and Applications, In *J. Shanahan, Y. Qu, and J. Wiebe (eds.)*, vol. 20, pp. 1–9.
- Saima Aman and Stan Szpakowicz. 2007. Identifying Expressions of Emotion in Text. *V. Matoušek and P. Mautner (Eds.): TSD 2007*, LNAI 4629, 196–205.
- Sebastiani Fabrizio. 2002. Machine learning in automated text categorization. *ACM Computing Surveys*, vol. 34(1).
- Strapparava Carlo and Mihalcea Rada. 2007. SemEval-2007 Task 14: Affective Text, *ACL*.
- Strapparava Carlo and Alessandro Valitutti. 2004. WordNet-Affect: an affective extension of WordNet. *4th LRE*, 1083-1086.
- Turney Peter D. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *ACL*, pp. 417- 424.
- Wilson Theresa, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *HLT/EMNLP 2005*.
- Zhang Yu, Li Zhuoming, Ren Fuji and Kuroiwa Shingo. 2008. A preliminary research of Chinese emotion classification model. *IJCSNS*, 8(11), 127-132.