Sentence Level Emotion Tagging on Blog and News Corpora

Dipankar Das and Sivaji Bandyopadhyay

Department of Computer Science and Engineering, Jadavpur University, Kolkata-700032, India

ABSTRACT: Here we report a sentence level emotion tagging system based on the word level constituents in Bengali blog and English news corpora. An emotion-annotated Bengali blog corpus was prepared manually with Ekman's six emotion tags using the emotion annotated English news headline corpus from SemEval 2007. The word level annotation is carried out semi-automatically. The baseline system at word level for each emotion class assigns the class label to each word. The Conditional Random Field (CRF) based classifier used for word level emotion tagging outperformed the baseline for each emotion class. Sentence-level emotion scores for each emotion class are calculated as the average word level emotion scores based on the SentiWordNet. The emotion tag with the highest score is assigned to the sentence, followed by a rule based post-processing technique for handling negative words. The system demonstrated the highest overall average 65% *F-Score* value for Bengali and 63.26% for English.

KEYWORDS: SentiWordNet, Blog, News, Emotion, CRF, Emotion tag

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 Analysis has a rich set of applications, ranging from tracking users' emotion about products or events or about

Correspondence: Prof. Sivaji Bandyopadhyay, 188, Raja S C Mallik Road, Jadavpur University, Kolkata-700 032, India; e-mail: sivaji_cse_ju@yahoo.com; dipankar.dipnil2005@gmail.com

politics as expressed in online forums, to customer relationship management. Question Answering (QA) systems and modern Information Retrieval (IR) systems are increasingly incorporating emotion analysis within their scope. Not only QA and IR but also a wide range of other Natural Language Processing tasks has started to use emotional information.

In addition, text does not only contain informative objective contents but also attitudinal private and subjective information, including emotional states. The objective of an emotion analysis system is to identify and classify any emotional text into some predefined emotional category (e.g. *happy*, *sad*, *anger*, *disgust*, *fear*, *surprise*, *love* etc.). Human emotion described in texts is an important cue for our communication, but the identification of emotional state from texts is hard because it is not open to any objective observation or verification.

Emotions may be expressed by a single word or a group of words. The sentence level emotion identification process plays an important role in tracking emotions or to discover the cues for generating such emotions or to properly identify them. Sentences are the basic information units of any document. The overall document level emotion identification process depends on the emotion expressed by the individual sentences of that document, which in turn, is based on the emotions expressed by the individual words. Obviously, some pragmatic and discourse-level analyses (e.g. idioms, metaphors etc.) are also necessary to track the actual context of the different emotional shades inscribed in the text.

Different unsupervised, supervised, and semi supervised strategies were adopted for decades to identify and classify emotions. Recently, an increasing amount of research was devoted to investigate the ways of recognizing favorable and unfavorable sentiments toward specific subjects within natural language texts (Pang et al. 2002). The characterization of words and phrases according to their emotive tone (Turney, 2002) and classification of reviews into *recommended* (thumbs up) and *not recommended* (thumbs down) using the *semantic orientation* of the phrases is one of the important tasks in this field. But in many domains of text, the values of individual phrases may bear little relation to the overall sentiment expressed by the text. Determining subjectivity and polarity measure, the work (Cardie et al. 2005) reports state-of-the-art results on Multi Perspectives Question Answering (MPQA) annotation.

The evaluation campaigns, TREC (Text REtrieval Conference) Blog tracks of successive years starting from 2006 and NTCIR (NII Test Collection for Information

Retrieval Systems) codenamed by National Institute of Informatics (NII) have drawn more and more attention towards the emotion related tasks (Ounis et al. 2008; Seki et al. 2006) during the last decade. Affective text shared task on news headlines at SemEval 2007 for emotion and valence level identification (Chaumartin, 2007) has drawn the focus to this field.

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). In addition, Blogs (Yang et al. 2007) are the communicative and informative repository of text based emotional contents in Web 2.0. Research on emotion shows that blogs play the role of a substrate to analyze the reactions of different emotional enzymes. Many blogs act as an online diary to the bloggers reporting daily activities and surroundings. Sometimes, blog posts are annotated by other bloggers. As the large blog data set is suitable for machine learning models, several supervised and unsupervised machine learning classification techniques on blog data for comparative evaluation (Mishne and Rijke 2006) are carried out in this area. Building emotion lexicons (Yang et al. 2007) using Yahoo! Kimo Blog corpus, emoticons were used to identify emotions associated with textual keywords and for classifying news articles according to the readers' emotions instead of the authors' is implemented in Lin et al. (2007). Multi-perspective question answering with preparation of subjectivity lexicon and opinion finder (Cardie et al. 2003; Stoyanov et al. 2005) views the task as of opinion-oriented information extraction. Opinions mining at word, sentence, and document levels from news and web blog articles with opinion summarization (Ku et al. 2006) were considered as topic- and genre-independent classification of the blog post into *objective*, *subjective-positive*, or *subjective-negative* classes.

Most works were carried out for English. Bengali is less privileged and less computerized than English. Works on emotion analysis in Bengali have begun recently (Das & Bandyopadhyay, 2009a). Most published machine learning-based models have considered the sentence as their basic key constituent, whereas the present work also deals with words for fine grained analysis. The comparative evaluation of the features on the equivalent domain for Bengali and English language can be found in (Das & Bandyopadhyay, 2009b).

In this paper, the investigation mainly focused on analyzing emotion in Bengali blog texts and news headlines of the English SemEval 2007 affect sensing corpus (Strapparava & Mihalcea, 2007) according to Ekman's (1993) six basic emotion types, such as *happiness, sadness, anger, fear, surprise,* and *disgust.* The word level annotation of both corpora was carried out semi automatically. The word level annotated data was verified by ourselves and found to be satisfactory on both blog and news data. The inter annotator agreement was carried out. The emoticons that appear in the Bengali blog texts were assigned with their proper emotion tags using a predefined domain dependent knowledge base. We used the Conditional Random Field (CRF) based machine learning approach (McCallum et al. 2001) for word level emotion tagging. Various experiments regarding symbolic feature, language, and domain-dependent features were carried out. The lexical feature (e.g. POS, SentiWordNet emotion word) outperforms other features significantly, whereas a different combination of context features shows significant improvement in performance. The word level tagging system has demonstrated F-Score values of 64.33% for Bengali and 72.27% for English on 2,500 word tokens of the development sets. Error analysis was incorporated and equal distribution of emotion tags with the non-emotion tag organized to improve word level emotion tagging. 70.23% and 83.65% overall F-Score values achieved on 1,500 word tokens of blog and news, respectively.

As no emotion-annotated corpus is available in Bengali, each test sentence of the Bengali blog corpus was annotated with single emotion tag and verified by us successively. The English annotated test sentences were separated into the emotion classes corresponding to the possible emotion types containing annotated scores in the news headlines. Evaluations for both blog and news were carried out separately for each of the six emotion classes. Six different emotion tag weights were calculated and applied on the word level emotion tagged test data to acquire sentence level emotion scores for each emotion type. A sentence level emotion tag that has the maximum emotion score was assigned to each sentence. The Bengali blog and the English news data, each containing 200 test sentences, showed the respective F-Scores of 63.98% and 66.66% for *happy,* 65.72% and 59.33% for *sad,* 60.66% and 62.32% for *anger,* 68.98% and 62.70% for *disgust,* 64.23% and 65.89% for *fear,* and 66.45% and 62.67% for *surprise* emotion classes.

The rest of the paper is organized as follows. The preparation of relevant resources is described in section 2. The baseline and the CRF-based model for word level emotion tagging system and its evaluation are specified in section 3. Section 4 describes the method of calculating six emotion tag-weights for assigning sentence level emotion tag and post processing steps for handling negative words. Emotion class-wise evaluation mechanisms and associated results are discussed in section 5. Finally, section 6 concludes the paper.

2. RESOURCE PREPARATION

The sentiment lexicon, *SentiWordNet* (Esuli & Sebastiani 2006) and emotion word lists like *WordNet Affect lists* (Strapparava & Valitutti 2004) are available in English, and these resources were used for emotion analysis task for English news data. Bengali is a resource poor language, however, and no emotion lexicon exists in Bengali. To overcome this problem, a Bengali *SentiWordNet* is being developed by replacing each word entry in the synonymous set of the English *SentiWordNet* by its possible set of Bengali synsets using a synset based English to Bengali bilingual dictionary that is being developed as part of the EILMT project[1]. Some words in the English *SentiWordNet* may not find their Bengali equivalent synsets, and such English words are transferred to the equivalent Bengali *SentiWordNet.* The modified Bengali *SentiWordNet* was considered as an emotion-based lexicon throughout the work on Bengali.

A knowledge base (shown in Table 1) for the emoticons was prepared after minutely analyzing the Bengali blog data. These emoticons are domain dependent. We classified the available emoticons into six emotion classes. Each image link of the emoticon in the raw corpus was mapped into its corresponding textual entity in the tagged corpus according to their proper emotion types using the knowledge base. This knowledge base was also used during the word level semi-automatic annotation of Bengali blog data.

The English *WordNet Affect* lists based on Ekman's six emotion types were updated with the synsets retrieved from the English *SentiWordNet* to make adequate number of emotion word entries. The updated information is shown in Table 2. These lists were converted to Bengali (Das & Bandyopadhyay 2009b) using the synset based English to Bengali bilingual dictionary as before. These six lists (for

[1] English to Indian Languages Machine Translation (EILMT) is a TDIL project undertaken by the consortium of different premier institutes and sponsored by MCIT, Govt. of India.

both English and Bengali) are termed as *Emotion lists* and were used during semi-automatic word level emotion annotation task.

The Bengali blog data were collected from the web blog archive (www.amarblog.com). The 14 different topics and their corresponding user comments containing 1500 sentences were retrieved. On the other hand, 1253 sentences, tagged with six emotion scores were retrieved from the English SemEval 2007 corpus (Strapparava & Mihalcea, 2007).

TABLE 1

Knowledgebase for tracking emoticons during annotation

Emoticons	Source Tag	Type
☺	<emo_icon_happy>	*happy*
☹	<emo_icon_sad>	*sad*
:-@	<emo_icon_ang>	*anger*
:-$	<emo_icon_dis>	*disgust*
:'(	<emo_icon_fear>	*fear*
:-O	<emo_icon_sur>	*surprise*
😐	<emo_icon_ntrl>	*neutral*

TABLE 2

Number of emotion word entries in the six *WordNet Affect lists* before and after the updating process for English and Bengali

WordNet Affect Lists	# No. of Emotion Word Entries		
	English		Bengali
	Before updating	After updating	After updating and converting by synset
Joy	538	4046	13572
Sadness	208	3796	6004
Anger	373	1379	5190
Disgust	71	347	2600
Fear	206	906	3754
Surprise	87	608	2308

3. WORD LEVEL EMOTION TAGGING

Primarily, the word level annotation was carried out semi-automatically. The assignment of an emotion tag to a word was done with the help of the *Emotion list* in which that word is present. Other non-emotional words were tagged with *neutral* type. We verified the word level emotion annotated data. One author verified the annotation of another. This annotation task including the inter-annotator agreement is being carried out. 8000 tokens of 12500 annotated Bengali blog word tokens and 7060 tokens of 11060 annotated English news word tokens were considered for training with the CRF based word level emotion tagging system. 2500 and 1500 word tokens from the rest of Bengali blog and English news respectively, were used as the development and the test sets. A baseline system for word level emotion tagging is implemented primarily to measure the performance of the system. Then a CRF based machine learning technique is used to evaluate the performance of the word level emotion tagging system over baseline. We observed that the CRF based word level emotion tagging system outperformed the baseline system.

3.1 Baseline Model

This module is implemented to identify the performance of word level emotion tagging for each emotion class with respect to all words from the six emotion classes without incorporating any prior knowledge regarding word features. Six separate modules were implemented for the six emotion classes corresponding to Ekman's six emotion types. The words are passed through these six separate modules for tagging each word with the emotion tag that represents its own emotion class. The baseline evaluation is made for six emotion classes separately. The comparative F-Scores of the CRF based model with the baseline model for the six emotion classes on test sets for both Bengali and English are shown in Table 5.

3.2 CRF-Based Model

Here, we used the CRF classifier for classifying emotion and non-emotion words into their appropriate classes and tag them with emotion or *neutral* tags.

3.2.1. *Feature selection & training.* Feature plays a crucial rule in the CRF framework. By manually reviewing the Bengali blog data and English news corpus and their different language specific characteristics, 10 active features were selected heuristically for our classification task. Each feature is Boolean in nature.

- *POS information:* We are interested in the *verb, noun, adjective,* and *adverb* words as these are generally emotion informative constituents.
- In total, 1500 blog sentences were passed through a Bengali part of speech tagger (Ekbal & Bandyopadhyay 2008) based on the SVM framework. The POS tagger was developed with a tagset of 26 POS tags[2], defined for the Indian languages. The POS tagger has demonstrated an overall accuracy of approximately 90%. The 1253 sentences collected from the English SemEval 2007 Affect Sensing Corpus were POS tagged with an open source Stanford Maximum Entropy based POS tagger (Manning & Toutanova 2000). The best reported accuracy for the POS tagger on the Penn Treebank is 96.86% overall and 86.91% on previously unseen words.
- *SentiWordNet emotion word*: A word appearing in the SentiWordNet (English and Bengali) generally contains emotion. For word level emotion classification, disambiguating the emotion and non-emotion words properly is necessary. This feature helps the classifier to clearly define emotion and non-emotion words.
- *F*irst sentence in a topic: The first sentence of the topic generally contains emotion (Alm et al. 2005). Yet, during sentence-level emotion analysis, all English news headlines considered for this work are equally important. So, this feature was identified as less significant during feature level analysis on the English development set.
- *Reduplication*: The reduplicated words (e.g., *bhallo bhallo* [good good], *khokhono khokhono* [when when] etc.) in Bengali are most likely emotion words. English reduplicated words ([so so] etc.) were also taken into consideration as they are emotion words.
- *Question words*: The question words generally contribute to the emotion in a sentence. (e.g. *how, why, what* etc.)
- *Colloquial/Foreign words*: The colloquial words (e.g., *kshyama* [pardon] etc.) and foreign words (e.g. in Bengali, Thanks, *gossya* [anger] etc.) are highly rich with emotional contents. Yet, the separation of these words from the corpus is not an easy task. Hence, these words are to be emotion tagged manually. The nature of their appearance in a sentence, like use of a single colloquial/foreign word to form a sentence, is important during classification.

---

[2] http://shiva.iiit.ac.in/SPSAL2007/iiit_tagset_guidelines.pdf

- *Special punctuation symbols*: The symbols (e.g., !, ?, @ etc) appearing at the word/sentence level convey emotions. Such symbols are appropriately tagged. The number of occurrences of special punctuation symbols attached to a word is an important feature with respect to word level classification.
- *Quoted sentence*: Such sentences, especially remarks or direct speech, generally contain emotion.
- *Length of a Sentence*: Sentence length is a crucial factor for the emotion classification task. Detailed experimental results on the effect of the sentence length on the system F-Scores are shown in Table 3 with Experiment Ids (x) (a), (x) (b), and (x) (c).
- *Emoticons*: The emoticons generally contribute as much as real sentiment to the words that precede or follow it. The consecutive occurrence of such emoticons emphasizes the preceding word or sentence.

ক্ষ্যামা দাও! "তুমি ভালো লোক"
(*khyama*) (*dao*)! "(*tumi*) (*bhalo*) (lok)"
(Forgive)! "(you) (good) (person)"

TABLE 3

Frequencies of different features in the training and test corpus of Bengali blog and English news data

Feature Types	Frequency in the corpus	
	Blogs Training Test	News Training Test
Parts of Speech	2132000678	1647000532
Word in SentiWordNet	684000157	1157000175
First Sentence	96000013	753000200
Reduplication	18000007	5000000
Question Words	13000001	23000009
Coll. / Foreign Words	35000009	8000000
Special Symbols	650000430	170000080
Quoted Sentence	22000008	7000003
Length of Sentence (>=8 and <15)	1732000157	1689000160
Emoticons	87000033	0000000

TABLE 4

F-Scores on 2500 Bengali and English word tokens of the development set

Expt. Id	Feature Combinations	F-Score (in %)	
		Blog	News
(i)	**Part of speech** (POS)	48.21	60.02
(ii)	***SentiWordNet* emotion word**	43.15	56.79
(iii)	POS + ***SentiWordNet* emotion word**	50.30	66.70
(iv)	POS +*SentiWordNet* emotion word+ **First sentence in a topic (Blog only)**	52.14	66.05
(v)	POS + *SentiWordNet* emotion word + **Reduplication**	53.26	67.12
(vi)	POS + *SentiWordNet* emotion word+ Reduplication + **Question words**	54.40	68.50
(vii)	POS + *SentiWordNet* emotion word+ Reduplication + Question words + **Colloquial / Foreign words**	55.18	69.10
(viii)	POS + *SentiWordNet* emotion word+ Reduplication + Question words + Colloquial / Foreign words + **Special punctuation symbols**	58.50	72.10
(ix)	POS + *SentiWordNet* emotion word+ Reduplication + Question words + Colloquial / Foreign words + Special punctuation symbols + **Quoted sentence**	59.70	74.96
(x)	POS + *SentiWordNet* emotion word+ Reduplication + Question words + Colloquial / Foreign words + Special punctuation symbols + Quoted sentence + **Length of a Sentence** Length of a Sentence (>15,>16) Length of a Sentence (<8,<7,<6) Length of a Sentence (>=8,<15)	56.23 52.80 61.03	69.33 65.76 76.80
(xi)	POS + *SentiWordNet* emotion word+ Reduplication + Question words + Colloquial / Foreign words + Special punctuation symbols + Quoted sentence + Length of a Sentence (>=8, <15) + **Emoticons (Blog only)**	62.03	76.80
(xii)	POS + *SentiWordNet* emotion word+ Reduplication + Question words + Colloquial / Foreign words + Special punctuation symbols + Quoted sentence + Length of a Sentence + Emoticons (**Blog only**) + **Context features** (unigram and bigram)	**64.33**	**78.56**

Different unigram and bi-gram context features (word level as well as POS tag level) and their combinations were generated from the training corpus. The above Bengali blog sentence contains four features (Colloquial word (*kshyama*), special symbol (!), quoted sentence and emotion word (আনন্দ [*happy*])) together and all these four features are important for identifying the emotion of this sentence.

3.2.2. *Evaluation.* The importance of different features varies according to their frequencies (as shown in Table 3) in the corpus. 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 in word level emotion tagging were measured on the development set and the results are given in Table 4. We added each feature into the active feature list at a time if the inclusion of the feature along with the pre-selected features improved the F-Score of the word level emotion tagging on development data. These active features were satisfactory also on the test data. The best achievable F-Score values and corresponding feature combinations are listed in Table 4.

The POS feature (only *adjective, noun, verb* and *adverb* words) and *SentiWordNet* emotion word feature played an important role in this task. During the development phase, certain features (e.g. First sentence in a topic) played no role in the word level emotion tagging process for English news sentences. This phenomenon is specific for the SemEval 2007 data as all sentences in the corpus are news headlines and hence, are always first sentences in the corresponding text document. The inclusion of such a feature did not improve the F-Score of English word level emotion tagging system and was removed from the feature set.

Special Symbols and their number of occurrences were selected as one of the features, which enhanced the performance of the system by around 3% for blog and news data, respectively. Detailed experimental results on the effect of the sentence length on the system F-Scores are shown in Table 4 with Experiment Ids (x) (a), (x) (b) and (x) (c). As the news headlines do not contain emoticons, this feature was therefore excluded from the active feature set for English. During the CRF-based training phase, the current token word with two previous and next words and their corresponding POS were selected as context feature for that word.

Evaluation results of the development set of 2500 word tokens demonstrated F-Score values of 64.33% for Bengali blog data and 78.56% for English news data. The comparative results (as shown in Table 5) of the baseline system with the CRF based system on the development set show that the CRF-based machine learning approach significantly outperformed the baseline model in the instance of word level emotion classification.

Error analysis was conducted with the help of confusion matrices as shown in Table 6 and Table 7 for Bengali blog and English news headlines respectively. A close investigation of the evaluation results suggests that the errors in CRF emotion tagging are primarily due to the uneven distribution between emotion and non-emotion tags. The number of non-emotional or neutral type tags is comparatively higher than that of other emotional tags in a sentence.

TABLE 5

F-Score (in %) values on the test set of Bengali and English baseline and CRF models

Emoticon Classes	Bengali		English	
	Baseline	CRF	Baseline	CRF
happy	57.12	62.45	65.25	78.43
sad	53.67	64.32	63.19	73.39
anger	52.50	60.08	70.20	76.24
disgust	53.11	68.76	65.38	77.07
fear	57.82	66.07	80.93	85.91
surprise	53.21	64.30	70.56	80.32

TABLE 6

Confusion matrix for the development set of Bengali Blog

Tag	*happy*0*sad*000*ang*00*dis*00*fear*000*sur*00*ntrl*
happy	00–0000.01000.05000.0000.00000.0000.03
sad	0.00600 –0000.02000.0300.00000.0000.02
ang	0.000000.03000–0000.0000.02000.0000.01
dis	0.000000.00000.01000–000.01000.0000.01
fear	0.000000.00000.00000.0000–0000.0000.01
sur	0.020000.00700.00000.0000.00000–000.012
ntrl	0.000000.00000.00000.0000.00000.0000–

TABLE 7

Confusion matrix for the development set of English News headlines

Tag	*Happy0sad00ang000dis000fear00sur00ntrl*
happy	0–00000.01000.03000.00000.00000.0000.02
sad	0.002000–0000.02000.03000.00000.0000.01
ang	0.010000.03000–0000.00000.01000.0000.01
dis	0.000000.00000.01000–0000.01000.0000.01
fear	0.000000.00000.00000.00000–0000.0000.01
sur	0.020000.00300.00000.00000.00100–000.012
ntrl	0.000000.00000.00000.00000.00000.0

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 Є *POS*. We have twenty six (26) subclasses in the Bengali POS tagset and 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 showed the improved F-Scores of 70.23% for blog and 83.65% for news with the inclusion of this class splitting technique applied on 1500 test word tokens. As Bengali is a morphologically rich language and takes variety of suffixes, the word level emotion tagging F-Score that was achieved is less in comparison with English.

4. SENTENCE LEVEL EMOTION TAGGING

This module was developed to identify sentence level emotion tags based on the word level emotion tags (assigned by the CRF based Emotion tagger) for each sentence of test set for each of the languages. Before that, we calculated sense based tag weights for each of the six emotion tags.

Sense_Tag_Weight (STW): This weight was calculated using SentiWordNet. We selected the basic six words *"happy"*, *"sad"*, *"anger"*, *"disgust"*, *"fear"* *"surprise"* as the seed words corresponding to each type of emotion tag. The positive and negative scores in the English *SentiWordNet* for each synset in which each of the seed words appear were retrieved and the average of the scores was fixed as the *Sense_Tag_Weight* of that particular emotion tag. Table 8 shows the values of **STW** of six emotion tags. Neutral tag was assigned with **STW** as zero.

TABLE 8

Sense_Tag_Weight for each of six emotion tags

Tag Type	***Sense_Tag_Weight***
emo_happy	0.0125
emo_sad	(-) 0.1022
emo_ang	(-) 0.5
emo_dis	(-) 0.075
emo_fear	0.0131
emo_sur	0.0625
emo_ntrl	0.0

Each sentence is assigned a *Sense_Weight_Score* (SWS) for each emotion tag which is calculated by dividing the total *Sense_Tag_Weight* (STW) of all occurrences of the emotion tag in the sentence by the total *Sense_Tag_Weight* (STW) of all types of emotion tags present in that sentence.

Thus, **SWSi = (STWi * Ni) / (∑ j=1 to 7 STWj * Nj)** | i Є j where **SWSi** is the Sentence level *Sense_Weight_Score* for the emotion tag i in the sentence and Ni is the number of occurrences of that emotion tag in the sentence. Each sentence was assigned with the sentence level emotion tag **SET** for which **SWSi** is highest. **SET** = [**max** i=1 to 7(**SWSi**)]. The sentences were tagged as neutral type if for all emotion tags i, **SWSi** produced zero (0) emotion score.

4.1 Post-Processing Strategies

The presence of negative words and their number of occurrences is significant in assigning the final emotion tag for a sentence. We implemented a rule-based post-processing module for handling negative words.

The consecutive occurrence of negative words does not reverse the assigned emotion type, whereas presence of a single negative word changes the actual emotion type in the completely opposite direction. For example, the following English News sentence:

Paris Journal: Smoking No Longer Tres Chic in France. (English News)

তুমি দুঃখ কোরো না। (Bengali Blog)
(*tumi*) (*dookkho*) (*koro*) (*na*)
you worry do not

was tagged as "*sad*" by the system but in the gold standard SemEval 2007 news corpus, the emotion scores for "*happy*" and "*surprise*" were assigned for this sentence. Considering the single occurrence of the negative word "NO" in the English sentence and "না" (*na*)(not) in the Bengali sentence during post processing phase, the emotion tag of both the sentences were reversed to "*happy*", the desired emotion tags for these sentences.

In the following sentence, two consecutive occurrences of negative words ("NO" and "NOT") do not change the actual emotion expressed by the sentence.

Seduced by Snacks? No, Not You. (English News)

In this case, the system has assigned the "*fear*" tag, which has the only significant emotion score in comparison with other emotion scores in the annotated English news corpus. We applied the rule that two consecutive negative words present in a sentence do not change the sentence level emotion tag. The same rule was applied for the following Bengali sentence and the annotated "*happy*" emotion tag has also been matched with the system generated '*happy*" emotion tag.

না না আমি এখন ভালো আছি। (Bengali Blog)
(*na*) (*na*) (*ami*) (*ekhon*) (*bhalo*) (*aachi*)
No No I now well am

5. EMOTION CLASS WISE EVALUATION

Each Bengali blog sentence in the test set was annotated with a single emotion tag. The 200 test sentences are classified into six different emotion classes according to their annotated sentential emotion tag. The test sentences of each emotion class were passed through the system for assigning a single emotion tag. The system-generated emotion tag of each test sentence was compared against its annotated emotion tag. Then the emotion class wise *recall, precision*, and *F-Scores* were measured for each of the emotion classes separately.

Each news headlines in the SemEval 2007 affect-sensing corpus was annotated with an individual score for each of the six emotion types. No single emotion tag was assigned for a sentence. The annotated emotion scores are in the range of zero to hundred [0-100] in the English news corpus. We extracted 200 test sentences that have some emotion scores for any emotion tag and each sentence is tagged with its corresponding emotion tags. The 200 test sentences were classified into six emotion classes. Of the 200 test sentences, 120 sentences are classified into more than one emotion class as these sentences are assigned with more than one type of emotion scores in the annotated English news SemEval 2007 affect-sensing corpus. We present the test set F-Scores for both the Bengali blog data and the English news data for each emotion class in Table 9.

TABLE 9

Test set *Recall, Precision* and *F-scores* out of total test sentences per emotion class for Blogs and News (in %)

Emoticon Classes*	Ref	Bengali Blogs			English News		
		Recall	*Precision*	*F-Score*	*Recall*	*Precision*	*F-Score*
Happy	[40, 33]	70.45	58.59	63.98	72.34	61.80	66.66
Sad	[31, 51]	76.89	57.38	65.72	60.23	58.45	59.33
Anger	[42, 56]	73.33	51.72	60.66	64.83	59.99	62.32
Disgust	[31, 43]	70.46	67.56	68.98	64.55	60.95	62.70
Fear	[33,48]	71.98	57.98	64.23	67.80	64.08	65.89
Surprise	[23, 39]	74.21	60.15	66.45	68.22	57.95	62.67

*[# Bengali sentences, # English sentences]

The results show that the system performed satisfactorily for both languages although there is a scope for improving the F-Score values. The recall of the system is comparatively higher than that of precision. The loss in precision occurred due to the frequent use of metaphoric words in blogs and news, as the metaphors are hard to tag with their emotional senses. The domains (blog and news) are distinct and not equivalent, but the system has demonstrated an overall satisfactory performance.

6. CONCLUSION

Emotion classification is a recent sub discipline at the crossroads of information retrieval and computational linguistics. An emotion tagging system for Bengali blog data and English news data that works at the word and the sentence level has been described in this work. We did not study the handling of metaphors and their impact in detecting sentence level emotion. The phrase level analysis of the input text is the future demand of this system to cope with context level discrepancies and fine tuned negation handling. The system can be used in an emotion-based information retrieval system in which the retrieved documents will match the user defined query word(s) and emotion specification. Sometimes, users of the blogs comment on others comments. The identification of such overlapped comments on a given topic is crucial for detecting emotion. Also the tracking of a single user's comments on the same topic, as well as on different topics, is really important to manage. These works along with document level analysis are the future areas to be explored.

REFERENCES

Alm Cecilia Ovesdotter, Roth Dan and Sproat Richard. 2005. Emotions from text: machine learning for text-based emotion prediction, *HLT-EMNLP*, 579-586.

Cardie Claire, Wiebe Janyce and Wilson Theresa. 2005. Annotating expressions of opinions and emotions in language, *Language Resources and Evaluation*, **39** (2), 165-210.

Cardie Claire, Wiebe Janyce, Wilson Theresa, Litman Diane J. 2003. Combining low-level and summary representations of opinions for multi-perspective question answering, *New Directions in Question Answering 2003*, 20-27.

Chaumartin François-Régis. 2007. UPAR7: A knowledge-based system for headline sentiment tagging, *SemEval-2007, ACL*, 422-425.

Das D. and Bandyopadhyay S. 2009a. Word to sentence level emotion tagging for Bengali blogs, *ACL-IJCNLP 2009*, 149-152. Suntec, Singapore.

Das D. and Bandyopadhyay S. 2009b. Emotion tagging – a comparative study on Bengali and English blogs, *ICON 2009*, Hyderabad, India.

Ekbal A. and Bandyopadhyay S. 2008. Web-based Bengali news corpus for lexicon development and POS tagging, *POLIBITS, ISSN 1870-9044*, **37**, 20-29.

Ekman Paul. 1993. Facial expression and emotion, *American Psychologist*, **48**(4), 384-392.

Esuli Andrea and Sebastiani Fabrizio. 2006. SENTIWORDNET: A publicly available lexical resource for opinion mining, *LREC-06.*

Ku Lun-Wei, Liang Yu-Ting, and Chen Hsin-Hsi. 2006. Opinion extraction, summarization and tracking in news and blog corpora, *AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs, AAAI Technical Report*, 100-107.

Lin K. H.-Y., Yang C. and Chen H.-H. 2007. What emotions news articles trigger in their readers? *SIGIR*, 733-734.

Manning Christopher D. and Toutanova Kristina. 2000. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger, *SIGDAT, EMNLP/VLC.*

McCallum Andrew, Pereira Fernando and Lafferty John. 2001. Conditional random fields: probabilistic models for segmenting and labeling sequence data, *ISBN*, 282-289.

Mishne G. and Rijke M. de. 2006. Capturing global mood levels using blog posts, *AAAI*, 145-152.

Ounis Iadh, Macdonald Craig and Soboroff Ian. 2008. Overview of the TREC2008 blog track, *TREC-2008.*

Pang Bo, Lee Lillian and Vaithyanathan Shivakumar. 2002. Thumbs up? Sentiment classification using machine learning techniques, *EMNLP*, 79-86.

Sebastiani Fabrizio. 2002. Machine learning in automated text categorization, *ACM Computing Surveys*, **34**(1).

Seki Yohei, Evans David Kirk, Ku Lun-Wei, Chen Hsin-Hsi, Kando Noriko, Lin Chin-Yew. 2006. Overview of opinion analysis pilot task at NTCIR-6, *NTCIR-6.*

Stoyanov Veselin, Cardie Claire, and Wiebe Janyce. 2005. Multi-Perspective Question Answering Using the OpQA Corpus, *HLT/EMNLP 2005*, Vancouver, Canada.

Strapparava Carlo and Mihalcea Rada. 2007. SemEval-2007 Task 14: Affective Text, *ACL.*

Strapparava Carlo and Valitutti Alessandro. 2004. WordNet-Affect: an affective extension of WordNet, *LRE*, 1083-1086.

Turney Peter D. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, *ACL*, 417- 424.

Yang C., Lin K. H.-Y., and Chen H.-H. 2007. Building emotion lexicon from weblog corpora, *ACL*, 133-136.

Yang Changhua, Lin Kevin Hsin-Yih and Chen Hsin-Hsi. 2007. Emotion classification using web blog corpora. *IEEE, WIC, ACM*, 275-278.

Zhang Yu, Li Zhuoming, Ren Fuji and Kuroiwa Shingo. 2008. A preliminary research of Chinese emotion classification model. *IJCSNS*, **8**(11), 127-132.