

Tracking Emotions of Bloggers – A Case Study for Bengali

Dipankar Das and Sivaji Bandyopadhyay

Abstract—The present paper describes the identification and tracking of bloggers' emotions with respect to time from the structured Bengali blog documents. The assignment of Ekman's six basic emotions to the bloggers' comments is carried out at sentence and paragraph level granularities. The Referential Informative Chain (*RIC*) developed for each blogger consists of the nodes representing the emotional states of that blogger. Each node of a *RIC* contains the identification information of its associated blogger, timestamp, section and emotional sentences. The nodes are arranged in each *RIC* based on the ascending order of the associated timestamps. An affect scoring technique has been employed to capture the emotions from each of the nodes of a blogger's *RIC*. The incorporation of self emotions and influential emotions as extracted from other bloggers plays a significant role in detecting the emotions of a blogger's present state. The *Extrinsic* evaluation produces *precision* (*P*), *recall* (*R*) and *F-Measure* of 61.05%, 69.81% and 65.13% respectively for evaluating the total of 193 emotional states of 20 bloggers. The *Intrinsic* evaluation has been conducted using a manual rater with the help of a statistical agreement coefficient, Krippendorff's alpha α . Two types of alpha, namely nominal alpha and interval alpha produce the average scores of 0.67 and 0.72, respectively.

Index Terms—Tracking, emotions, bloggers, affect score, agreement.

I. INTRODUCTION

SENTIMENT Analysis and Opinion Mining have been attempted with more focused perspectives rather than fine-grained emotions [1]. 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 [2]. The determination of emotions expressed in the text with respect to reader or writer is itself a challenging issue [3] as emotion is not open to any objective observation or verification [4]. Not only the classification of reviews [5], newspaper articles [6] or blogs [7], a wide range of other Natural Language Processing (NLP) tasks such as tracking users' emotion about products or events or about politics as expressed in online forums or news, to customer relationship management are also using emotional information.

Researches on emotion show that blogs play the role of a substrate to analyze the reactions of different emotional

enzymes. Many blogs act as online diaries of the bloggers reporting their daily activities and surroundings. Sometimes, the blog posts are annotated by other bloggers. If we consider such bloggers as the emotion holders of different stance on diverse social and political issues, we can have a better understanding of the relationships among countries or among organizations [8].

Among all concerns, sentiments or emotions of people are important because people's sentiment has great influence on our society. Recently, the identification of the temporal trends of sentiments and topics has drawn attention of NLP communities [9]. The perspectives of sociology, psychology and commerce along with the close association among people, topic and sentiment motivate us to investigate the insides of emotional changes of people over topic and time.

The present task involves the identification and tracking of bloggers' emotions with respect to time from Bengali web blog archive¹. The sections of the bloggers' comments for a given topic contain nested tree like structures along with distinguishable information regarding individual blogger and his/her associated blog posting timestamps. Presently, the comment sentences are already annotated during the development of the Bengali emotional blog corpus [10]. Hence, by default, each of the comment sections is assigned with the types of emotions that are present in the annotated sentences of that section. A Referential Informative Chain (*RIC*) for each blogger is developed by using the default annotated timestamp, unique identifier and emotional comments. These information are acquired from the nested tree like comment sections of each blog document. Each node of a *RIC* denotes the emotional state of its corresponding blogger at a particular time instance and all nodes of a *RIC* are arranged based on the ascending order of the associated timestamps.

The identification of Ekman's [23] six basic emotions from the bloggers' comments is carried out at sentence and paragraph level granularities using the information of word level constituents along with sense based scoring mechanism [11]. An affect scoring technique has been employed to identify the emotions of a state or node in each of the Referential Informative Chains (*RICs*) of the bloggers. The *Self Affect Score* (*SAS*) is added with the *Influential Affect Score* (*IAS*) to produce the *Emotional Score* (*ES*) of a blogger at each node with a particular timestamp. We have considered the *Self Affect Score* (*SAS*) of a blogger at a particular

Manuscript received on November 12, 2010, accepted on December 2, 2010.

D. Das is with the Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, 700032 (phone: +91-9432226464; e-mail: dipankar.dipnil2005@gmail.com).

S. Bandyopadhyay is with Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, 700032 (phone: +91-9433579595; e-mail: sivaji_cse_ju@yahoo.com).

¹ www.amarblog.com

timestamp as the *Emotion Score (ES)* of that blogger at its immediately previous timestamp. The *Influential Affect Score (IAS)* is the cumulative summation of the *Emotion Scores (ES)* of all other participated bloggers who commented between the blogger's immediate previous timestamp to its current timestamp. In case of calculating the *Influential Affect Score (IAS)* of a blogger, we have hypothesized that the *Emotion Scores (ES)* of other participated bloggers are to be calculated independently without considering the inclusion of the self and influential affects of those bloggers.

The best three Ekman's emotions that are acquired from the ordered *Emotion Scores (ESs)* are assigned to the associated node of the blogger's corresponding *RIC* as the identified emotions. The change of a blogger's emotions has been tracked based on the emotions that are assigned to the nodes of the blogger's *RIC*.

The importance of self and influential affects in tracking results has been evaluated in two stages based on the identified emotions at each node. One is the *Extrinsic* evaluation that is carried out through the standard metrics such as *precision (P)*, *recall (R)* and F-Measure and another one is the *Intrinsic* evaluation that considers the involvement of a manual rater and measures the performance using the statistical agreement coefficient, Krippendorff's alpha α [21] [22]. The system achieves *precision (P)*, *recall (R)* and F-Measure of 61.05%, 69.81% and 65.13% respectively in case of *Extrinsic* evaluation. On the other hand, two types of alpha, nominal alpha ($N\alpha$) and interval alpha ($I\alpha$) produces the satisfactory average scores of 0.67 and 0.71 respectively for evaluating a total number of 193 emotional states of 20 bloggers from a collection of 34 blog documents.

The rest of the paper is organized as follows. Section II describes the related work. The preparation of Referential Informative Chains (*RICs*) from the annotated and nested comment sections of the structured blog documents is discussed in Section III. Automatic emotion tagging of the bloggers' comments is described in Section IV. The affect scoring technique for identifying the emotion states and their tracking are discussed in Section V. The evaluation schemes along with the performance measures of the emotional states are discussed in Section VI. Finally, Section VII concludes the paper.

II. RELATED WORK

The earlier researches on emotion analysis reveal that a large number of contributory attempts were conducted for identification, generation or classification of emotions rather than tracking of emotions over event, topic or time. Mishne and de Rijke [7] proposed a system, *MoodViews*² to analyze the temporal change of sentiment. *MoodViews* analyzes multiple sentiments by using 132 sentiments used in LiveJournal³. Although our concept for the emotion tracking is similar to *MoodViews*, instead of using only the temporal

relations, we focus on identifying the changes of emotions of the blog users over time. Another system called *ThemeRiver* [15] visualizes thematic flows along with timeline. The present approach is different from *ThemeRiver* as we focus on tracking the emotion flows of the bloggers considering their self as well as influential emotions. The temporal sentiment identification from social events has been carried out in [9]. In their task, the authors have analyzed the temporal trends of sentiments and topics from a text archive that has timestamps in weblog and news articles and produces two kinds of graphs, *topic graph* that shows temporal change of the topics associated with a sentiment, and *sentiment graph* that shows temporal change of sentiments associated with a topic. In contrast, the present task considers the involvement of self and influential emotions in determining the emotions of the bloggers at a particular timestamp or at each emotional state. We have constructed Referential Informative Chain (*RIC*) for individual blogger to track emotions associated with each node of its *RIC*.

III. REFERENTIAL INFORMATIVE CHAIN (RIC) PREPARATION

The mode of language technology has changed dramatically since the last few years with the web being used as a data source in a wide range of research activities. There is a long history of creating a standard for western language resources. The Human Language Technology (HLT) society in Europe has been particularly zealous for the standardization of European languages [16]. The authors also mentioned that, in spite of having great linguistic and cultural diversities, Asian language resources have received much less attention than their western counterparts. In Asia, India is a multilingual country with a diverse cultural heritage. Indian languages are resource constrained. Bengali is the fifth popular language in the World, second in India and the national language in Bangladesh but it is less privileged and less computerized compared to English. Following observations have motivated us to develop the Bengali emotion corpus from the web.

Recent study shows that non-native English speakers support the growing use of the Internet⁴. The focus is to improve the multilingual search engines on the basis of sentiment or emotion. This raises the demand of linguistic resources for the languages other than English.

Majority of the existing works in this field have been conducted for English [17]. To the best of our knowledge, at present, there is no such available corpus that is annotated with detailed linguistic expressions for emotion in Bengali or even for other Indian languages. Hence, the work is a foray into emotion analysis for an Indian language such as Bengali.

The blog documents are stored in the format as shown in Figure 1 after retrieval from the Bengali web blog archive (www.amarblog.com). Each of the blog documents is assigned with a unique identifier (*docid#*) followed by a section devoted for topic and several sections devoted for different users' comments. Each comment section consists of several nested

² <http://moodviews.com/>

³ <http://www.livejournal.com/>

and overlapped sub sections that also contain the bloggers' comments. Each of the comment sections of an individual blogger is uniquely identified by the notion of section identification number (*secid#*). Each section contains the information regarding identification number of the blog user (*uid#*) and associated timestamp (*tid#*).

We have considered the individual comment section as separate paragraph that contains several emotional sentences. The sentences that are present in the bloggers' comment sections are already annotated during the development of Bengali emotional blog corpus [10]. Hence, by default, each of the comment sections is assigned with all types of emotions that are present at the annotated sentences of that section. The emotions present in such individual comment section represent the emotional state of the blogger at that timestamp. The annotated emotional states are verified by the authors. It is observed that the annotation task for sentential emotions compensate the manual effort of verifying the emotional states and the result produces satisfactory impression.

The Referential Informative Chain (*RIC*) for each of the bloggers is constructed by acquiring the default annotated information like timestamp (*tid#*), unique identifier (*uid#*) and emotional comments that are acquired from the nested tree like structure of the comment sections. Though all of the comment sections (as shown using the tag `<User Comments id=UC#>`) in the individual blog documents refer to a single topic, presently we have considered each comment section as a separate unit by assuming that no inter emotional impact exists among the various comment sections of a single blog document.

The individual *RIC* is developed for each single blogger with respect to each comment section. Each node of a *RIC* denotes the emotional state of the blogger at a particular time instance and the sequence of adding information into the nodes is based on the ascending order of associated timestamps. For example, in Figure 1, the two nodes namely *n1* and *n2* will be added into the front of the *RIC* developed for the blog user with *uid*=1. The associated timestamps (*t1*, *t4*) and emotions will also be added into the nodes accordingly. As *t4* > *t1*, the inclusion of node *n1* is considered before the inclusion of *n2* into the corresponding *RIC* of *uid* 1.

IV. AUTOMATIC EMOTION TAGGING

Each of the comment sections has been considered as a separate paragraph that contains the emotional sentences. The hypothesis of defining an emotional state with respect to each comment section or paragraph is to include all of the annotated sentential emotions that are contained in that paragraph. It is said that sentiment is typically a localized phenomenon that is more appropriately computed at the paragraph, sentence or entity level [24]. Hence, our primary investigation mainly aims to automatically identify the emotions at sentence and paragraph level from the bloggers' comment sections. Ekman's (1993) six basic emotion types, such as *happiness*, *sadness*, *anger*, *fear*, *surprise*, and *disgust* are considered to

accomplish the emotion tagging task as these emotions are termed as universal emotions [23]. The non emotional sentences are considered as neutral.

```

<DOC docid = xy>
  +<Topic>.... </Topic>
  -<User Comments id=UC1>
    -<U uid=1, tid=t1, secid=UC1>...
    -<U uid=2, tid=t2, secid=UC1.1>...</U>
    -<U uid=3, tid=t3, secid=UC1.2>...</U>
    -<U uid=1, tid=t4, secid=UC1.2.1>...</U>
    ....
  </U>
</User Comments>
+<User Comments id=UC2>
+<User Comments id=UC3>
...
</DOC>

```

Fig. 1. General Structure of a blog document

We have employed a sentential emotion tagging system [11] that consists of two prong approach, word level followed by sentence level. The Conditional Random Field (*CRF*) [19] based machine learning approach that incorporates several singleton features (e.g. *Part of Speech (POS) of the words*, *Question words*, *Reduplication*, *Colloquial / Foreign words*, *Special Punctuation Symbols*, *Negative words*, *Words of Quoted sentence*, *Emoticons*), context features (*unigram*, *bigram*) at word level as well as POS tag level along with different combinations of singleton and context features has been used for word level emotion tagging. The only difference that has been considered in our present attempt is the use of the Bengali *WordNet Affect Lists* [13] instead of using the Bengali *SentiWordNet*. The incorporation of error analysis and equal distribution of emotion tags with the non-emotion tag improves the word level emotion tagging system. The system demonstrates satisfactory performance with an average accuracy of 66.74% with respect to all emotion classes [18].

The sense based and corpus based scoring strategies are applied on the acquired word level emotion constituents to identify the sentence level emotion tags. The corpus based scores are calculated based on the frequency of occurrence of an emotion tag with respect to the total number of occurrences of all six types of emotion tags in an annotated corpus whereas sense based scores are calculated using *SentiWordNet* [20]. But, we have considered only the sense based scoring technique to accomplish our present research goal. The *Sense_Tag_Weights (STWs)* are the *tag weights* that are calculated using English *SentiWordNet* [20]. The basic six words “*happy*”, “*sad*”, “*anger*”, “*disgust*”, “*fear*” “*surprise*” are selected as the *seed words* corresponding to each emotion type. The *positive* and *negative* scores of each synset in which each of these *seed words* appear are retrieved from the English *SentiWordNet* [20] and the average of the scores is fixed as the *Sense_Tag_Weight (STW)* of that particular emotion tag (*happy*: 0.0125, *sad*: - 0.1022, *anger*: - 0.5, *disgust*: - 0.075, *fear*: 0.0131, *surprise*: 0.0625, and *neutral*: 0.0) [11].

Each sentence is assigned with a *Sense_Weight_Score (SWS)* for each emotion tag which is calculated by dividing the total *Sense_Tag_Weights (STWs)* of all occurrences of an

⁴ <http://www.internetworldstats.com/stats.htm>

emotion tag in the sentence by the total *Sense_Tag_Weights* (*STWs*) of all types of emotion tags present in that sentence. The *Sense_Weight_Score* is calculated as $SWS_i = (STW_i * N_i) / (\sum_{j=1 \text{ to } 7} STW_j * N_j) \mid i \in j$ where SWS_i is the sentence level *Sense_Weight_Score* for the emotion tag i and N_i is the number of occurrences of that emotion tag in the sentence. STW_i and STW_j are the *Sense_Tag_Weights* for the emotion tags i and j respectively.

On the other hand, the polarity information (i.e. *positive* and *negative*) that is associated with the word level *Sense_Tag_Weights* (*STWs*) is carried forward to the sentence level *Sense_Weight_Scores* (*SWSs*) for each emotion types. Therefore, we preserve the information of *Sense_Weight_Scores* (*SWSs*) to utilize them in calculating the *Affect_Scores* (*ASs*) of each comment section or paragraph. Though we have considered all the annotated sentential emotions as the default emotions of its corresponding paragraph, but the system produces the *Affect_Scores* (*ASs*) for individual emotional state by summing up the sentential *Sense_Weight_Scores* (*SWSs*) with respect to each of the six emotion types.

V. AFFECT SCORING

The *Affect_Scores* (*ASs*) for each of the six emotions are preserved to identify the emotional states of an individual blogger in its corresponding Referential Informative Chain (*RIC*). By assuming each of the comment sections as an interactive system with respect to bloggers' communication, in addition to *Affect_Score* (*AS*), two other types of *Affect_Scores* (*ASs*) are also used to determine the six *Emotion_Score* (*ES*) of each blogger who participated in that interactive system. The *Self_Affect_Scores* (*SASs*) are used to measure the effect of the blogger's own previous emotions in identifying its present emotional state whereas the *Influential_Affect_Score* (*IASs*) are considered for measuring the previous impact of all other bloggers' emotions in identifying a blogger's present emotional state. The two *Affect_Scores* (*ASs*), *Self_Affect_Score* (*SAS*) and *Influential_Affect_Score* (*IAS*) are calculated based on Ekman's six basic emotion classes. *Emotion_Score* (*ES*) for each of the six emotion classes is the summation of the three *Affect_Scores* (*ASs*) corresponding to that emotion class, $ES_i = AS_i + SAS_i + IAS_i \mid i \in \text{Ekman's six emotion classes}$.

If the target blogger is a beginner or blog starter or who first starts commenting, the *Emotion_Scores* (*ESs*) of the blogger are then the present *Affect_Scores* (*ASs*), i.e., $ES_{t0} = AS_{t0}$, where t_0 is the first timestamp associated with the first node of the target blogger's *RIC*. In other cases, *Self_Affect_Scores* (*SASs*) of a target blogger is the *Emotion_Scores* (*ESs*) at its immediate previous timestamp and the *Influential_Affect_Score* (*IAS*) is the cumulative summation of the *Emotion_Scores* (*ESs*) of all other participated bloggers who commented between the target blogger's immediate previous timestamp and present timestamp. The *Self_Affect_Scores* (*SASs*) as well as the *Influential_Affect_Scores* (*IASs*) for an initiator target blogger at the start timestamp are considered as zero. For a

non initiator target blogger, only the *Self_Affect_Scores* (*SASs*) are considered as zero at the start timestamp as it first starts commenting from that timestamp. The emotions of the target blogger are calculated at each timestamp by assuming that the emotions of other participating bloggers are not biased by other bloggers' emotions including the target till that timestamp. Hence, in case of calculating the *Influential_Affect_Score* (*IAS*) of a target blogger, we have hypothesized that the *Emotion_Scores* (*ES*) of other participating bloggers are to be calculated independently without considering the inclusion of the self and influential affects for these participating bloggers.

For example, if we consider the following snapshot from Figure 1 to calculate the *Emotion_Scores* (*ESs*) of the blogger with $uid=1$ at different timestamps or assign the *Emotion_Scores* (*ESs*) to the emotional states or nodes of its *RIC*, we need to consider both the *Self_Affect_Score* (*SAS*) as well as the *Influential_Affect_Score* (*IAS*). As this target blogger is an initiator of its corresponding comment section, at timestamp t_1 , the six *Emotion_Scores* (*ESs*) of the target blogger are equal to the six *Affect_Scores* (*AS*) that are calculated from the sentences of the corresponding comment section with $secid=UC1$. The first node of the target blogger's *RIC* is also assigned with the six *Emotion_Scores* (*ES*). At timestamp t_4 , the second node of the blogger's *RIC* is assigned with the six *Emotion_Scores* (*ESs*) that are calculated based on the *Self_Affect_Scores* (*SASs*) as well as the *Influential_Affect_Scores* (*IASs*). Not only the six *Affect_Scores* (*ASs*) that are calculated from the sentences of the comment section ($secid=UC1.2.1$) but the immediate previous *Emotion_Scores* (*ESs*) (in this case the *Affect_Scores* (*ASs*) of the target blogger's previous comment section $secid=UC1$) are also added as the *Self_Affect_Scores* (*SASs*) to calculate the six *Emotion_Scores* (*ESs*) at t_4 . The *Self_Affect_Scores* (*SASs*) are included in the present task to consider the emotional impact of a blogger's previous emotions into its present emotional state.

In addition to *Affect_Scores* (*ASs*) and *Self_Affect_Scores* (*SASs*), the six *Influential_Affect_Scores* (*IASs*) are also added. The *Influential_Affect_Scores* (*IASs*) are included to consider the emotional impact of other previous participating bloggers into the present emotional state. For each of the six emotion classes, the *Influential_Affect_Score* (*IAS*) is the cumulative summation of the *Emotion_Scores* (*ESs*) of all other participating bloggers between the timestamps t_1 and t_4 . In this case, the *Influential_Affect_Score* (*IAS*) for each of the six emotion types is calculated based on the *Emotion_Scores* (*ESs*) of other participating bloggers (with $uid=2$ and $uid=3$) at timestamp t_2 and t_3 (as $t_1 < t_2$, $t_3 < t_4$). The *Emotion_Score* (*ES*) of the blogger ($uid=1$) at timestamp t_4 with respect to a particular emotion class is defined as follows:

$$ES_{t4} = AS_{t4} + SAS_{t4} + IAS(t_1, t_4) \text{ where } SAS_{t4} = ES_{t1} \text{ and } IAS(t_1, t_4) = ES_{t2}(uid=2) + ES_{t3}(uid=3)$$

```
-<U uid=1, tid=t1, secid=UC1>...
-<U uid=2, tid=t2, secid=UC1.1>...</U>
-<U uid=3 tid=t3, secid=UC1.2>...</U>
-<U uid=1 tid=t4, secid=UC1.2.1>...</U>...
```

It has to be mentioned that, as *Influential Affect Score (IAS)* should have some emotional impact; therefore influence of emotions on other bloggers is being taken into consideration. Each of the nodes in the blogger's RIC is assigned with the six *Emotion Scores (ESs)*. The best three emotions based on the ordered *Emotion Scores (ESs)* of a node are tagged as the emotional state of the blogger at that timestamp. The tracking of a blogger's emotional states via the nodes of its corresponding RIC is carried out based on the associated the best three emotions. The change of a blogger's emotions has been tracked based on the emotions that are assigned to the nodes of the blogger's RIC.

Tracking along the time dimension is trivial because most text documents for analysis come with time when they were written, which is especially true for online documents such as product reviews, forum postings, and blogs. The challenge is still the accuracy of sentiment or emotion prediction and solving the associated problems [24]. Hence, the importance of self and influential affects in tracking results has been evaluated in two stages based on the identified emotions at each node.

VI. EVALUATION

The evaluation of the emotion tracking system has been carried out at two levels to measure the performance of the system. One is the *Extrinsic* evaluation that is carried out through the standard metrics such as *precision (P)*, *recall (R)* and *F-Measure* and the other is the *Intrinsic* evaluation that deals with respect to the agreement between manual annotation and system generated output. We have measured the Krippendorff's (2004) [21] [22] alpha coefficient, a standard metric used for inter-annotator reliability studies to consider the evaluator disagreements. The metric is also used in counseling and survey research, psychological testing, observational studies etc. The metrics for *Extrinsic* evaluation gives the coarse-grained statistics whereas the motivation of selecting the metric for *Intrinsic* evaluation is to identify the fine grained clarity that exists among the emotions present in the emotional states or nodes of an individual blogger's Referential Informative Chain (RIC).

The evaluation has been carried out with respect to a total of 193 emotional states of 20 bloggers from a collection of 34 blog documents. In case of *Extrinsic* evaluation, the system generated emotions are evaluated against the annotated emotions. We have not considered the emotions of individual states or nodes of the RICs. For the very reason, the *precision (P)*, *recall (R)* and *F-Measure* have been calculated with respect to all emotion classes. It is found that the incorporation of *Self Affect Score (SAS)* and *Influential Affect Score (IAS)* improve both *precision (P)*, *recall (R)* metrics. The results are shown in Table I.

Though the coarse grained evaluation shows satisfactory performance of the system, the fine grained clarity of assigned emotions in each of the states or nodes of the bloggers' RICs is evaluated by measuring the inter-rater agreement using

TABLE I
EXTRINSIC EVALUATION

Techniques	Precision (P)	Recall (R)	F-Measure
Affect Scoring (AS)	45.87	56.05	51.34
Self Affect Scoring (SAS)	16.22	28.12	22.10
Influential Affect Scoring (IAS)	20.63	32.98	26.72
AS+SAS	55.44	63.04	59.25
AS+IAS	59.66	65.35	62.82
AS+SAS+IAS	61.05	69.81	65.13

Krippendorff's alpha [21]. This evaluation technique is also termed as *Intrinsic* evaluation. It is a statistical measure of the agreement achieved when coding a set of units of analysis in terms of the values of a variable. The equation of alpha is as follows:

$$metric\alpha = 1 - \frac{D_o}{D_e} = 1 - \frac{\sum_{c=1, k=1}^v o_{ck} metric\delta_{ck}^2}{\frac{1}{n-1} \sum_{c=1, k=1}^v n_c n_k metric\delta_{ck}^2}$$

where D_o and D_e are the observed and expected disagreements, $metric\delta_{ck}^2$, a difference function between values c and k reflect the metric properties (Levels of Measurement) of the variable and O_{ck} , a coincidence matrix that cross tabulates the n pair-able values from the canonical form of the reliability data into a v -by- v square matrix, where v is the number of values available in a variable. The definition of n_c , n_k and n are as follows:

$$n_c = \sum_{k=1}^v o_{ck}, \quad n_k = \sum_{c=1}^v o_{ck}, \quad \text{and} \quad n = \sum_{c=1, k=1}^v o_{ck}$$

We have considered two metrics, *Nominal Alpha (N α)* and *Interval Alpha (I α)* for measuring the agreement. Hence, the following equations corresponding to their difference function has been considered to accomplish the goal:

$$nominal\delta_{ck}^2 = \begin{cases} 0 & \text{iff } c = k \\ 1 & \text{iff } c \neq k \end{cases} \quad \text{and} \quad interval\delta_{ck}^2 = (c - k)^2$$

On the other hand, the Krippendorff's alpha is applicable to any number of coders, each assigning one value to one unit of analysis, to incomplete (missing) data, to any number of values available for coding a variable, to binary, nominal, ordinal, interval, ratio, polar, and circular metrics (Levels of Measurement), and it adjusts itself to small sample sizes of the reliability data.

Thus, we have considered the emotional states of any blogger's RIC as the set of units to be analyzed. The four values out of which three correspond to the number of emotions assigned to the state and one value corresponds to the undetermined emotion. It has been observed that the system assigns the six *Emotion Scores (ESs)* of value zero to some nodes in the RICs. These states are termed as *Undetermined Emotional States (UESs)*. The evaluation technique considers the *Undetermined Emotional States (UESs)* as the incomplete or missing data. Otherwise, the best one or two or three emotions are assigned to the nodes based

on the availability of ordered *Emotion Scores (ESs)* with values greater than zero as obtained by the affect scoring technique.

The *Self Affect Scores (SASs)* and *Influential Affect Scores (IASs)* have been introduced to minimize the production of *Undetermined Emotional States (UESs)* assuming each current state of the blogger contains more or less affections from its previous emotional states. The requirement of normalizing the produced *Emotion Scores (ESs)* for each blogger was not faced as the average length of the *RICs* is not very large and the *Emotion Scores (ESs)* of other bloggers are independently considered during the calculation of *Influential Affect Scores (IASs)*. On the other hand, the identified emotions for each node of a *RIC* are stored in a vector. One manual rater has been appointed to carry out the assignment of the values to each unit based on the identified emotions and their correctness. The manual assignment has been performed against the annotated assignment. The annotated states also contain one or multiple or no emotions and similarly these gold standard emotions are represented using a vector. Each of the vectors for each node of a blogger's *RIC* is considered as a unit and any of the four values is assigned to each unit based on the number of emotions present in annotated vector and system generated vector.

Each of the nodes or emotional state of the bloggers' *RICs* is assigned the numeric values of the two raters. The *Nominal Alpha (N α)* and *Interval Alpha (I α)* are calculated for each of the bloggers' *RICs*. The number of emotional states or nodes of a blogger's *RIC* represents the number of units for analysis. We have shown the average scores in Table II with respect to a total of 193 emotional states of 20 bloggers. It is observed that the agreements have produced the alpha scores of 0.67 and 0.71 for *Nominal Alpha (N α)* and *Interval Alpha (I α)* respectively. $I\alpha > N\alpha$ signifies that the disagreements happen not largely among the neighboring values but the primary observation suggests that the incorporation of the two affect scores, *Self Affect Scores (SASs)* and *Influential Affect Scores (IASs)* substantially reduces the disagreement among the raters.

VII. CONCLUSION

In this paper, we have reported our work on identification and tracking of bloggers' emotions from structured Bengali blog documents. The present system shows the effectiveness of utilizing the previous information regarding self emotions and other bloggers' influential emotions in case of identifying present emotion of any blogger. Both of the *Extrinsic* and *Intrinsic* evaluations show significant improvement in the performance of the system. Our future plan is to employ the system in identifying emotional changes from lengthy referential chains to analyze the potential reasoning behind the change. The dependency among the bloggers' emotions is to be analyzed for capturing the reasons of emotional changes too. The hypothesis of the present model will be used in future for developing a topic driven emotion tracking model.

TABLE II
INTRINSIC EVALUATION

Techniques	Nominal Alpha (N α)	Interval Alpha (I α)
Affect Scoring (AS)	0.35	0.43
Self Affect Scoring (SAS)	0.15	0.24
Influential Affect Scoring (IAS)	0.23	0.31
AS+SAS	0.56	0.58
AS+IAS	0.59	0.62
AS+SAS+IAS	0.67	0.72

ACKNOWLEDGEMENTS

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

REFERENCES

- [1] C. Quan and F. Ren, "Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis," in *Empirical Method in Natural Language Processing – Association for Computational Linguistics*, pp. 1446-1454, 2009.
- [2] Y. Zhang, Z. Li, F. Ren and S. Kuroiwa, "A Preliminary Research of Chinese Emotion Classification Model," in *IJCSNS International Journal of Computer Science and Network Security*, vol. 8(11), pp. 127-132, 2008.
- [3] C. Yang, K. H. Y. Lin and H.H. Chen, "Writer Meets Reader: Emotion Analysis of Social Media from both the Writer's and Reader's Perspectives," in *009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, pp. 287-290, 2009.
- [4] R. Quirk, S. Greenbaum, G. Leech and J. Svartvik, *A comprehensive Grammar of the English Language*, Longman, New York, 1985.
- [5] P.D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," in *Annual Meeting of the Association for Computational Linguistics*, pp. 417-424, 2002.
- [6] K. H.-Y. Lin, C. Yang, and H.-H. Chen, "What Emotions News Articles Trigger in Their Readers?" in *SIGIR*, pp. 733-734, 2007.
- [7] G. Mishne and M. de Rijke, "MoodViews: Tools for Blog Mood Analysis," in *AAAI 2006 Spring Symposium on Computational Approaches to analyzing Weblogs*, 2006.
- [8] Soo-Min Kim and E. Hovy, "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text," in *ACL*, 2006.
- [9] T. Fukuhara, H. Nakagawa and T. Nishida, "Understanding Sentiment of People from News Articles: Temporal Sentiment Analysis of Social Events," in *ICWSM'2007*, Boulder, Colorado, USA, 2007.
- [10] D. Das and S. Bandyopadhyay, "Labeling Emotion in Bengali Blog Corpus – A Fine Grained Tagging at Sentence Level," in *8th Workshop on Asian Language Resources (ALR8), COLING*, pp. 47-55, 2010.
- [11] D. Das and S. Bandyopadhyay, "Word to Sentence Level Emotion Tagging for Bengali Blogs" in *ACL-IJCNLP*, pp. 149-152, 2009.
- [12] D. Das and S. Bandyopadhyay, "Sentence Level Emotion Tagging on Blog and News Corpora," *Journal of Intelligent System (JIS)*, vol. 19(2), pp. 125-134, 2010.
- [13] D. Das and S. Bandyopadhyay, "Developing Bengali WordNet Affect for Analyzing Emotion," in *International Conference on the Computer Processing of Oriental Languages*, pp. 35-40, 2010.
- [14] S. Sood and L. Vasserman, "ESSE: Exploring Mood on the Web," in *AAAI Conference on Weblogs and Social Media (ICWSM) Data Challenge Workshop*, 2009.
- [15] S. Havre, E. Hetzler, P. Whitney, and L. Nowell, "ThemeRiver: Visualizing Thematic Changes in Large Document Collections," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 8(1), pp. 9-20, 2002.

- [16] A. Ekbal and S. Bandyopadhyay, “A Web-based Bengali News Corpus for Named Entity Recognition,” in *Language Resources and Evaluation* vol. 42(2), pp. 173-182, 2008.
- [17] B. Pang and L. Lee, “Opinion Mining and Sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2(1-2), pp. 1-135, 2008.
- [18] D. Das and S. Bandyopadhyay, “Emotion Tagging – A Comparative Study on Bengali and English Blogs,” in *International Conference on Natural Language Processing*, pp. 177-184, 2009.
- [19] J. Lafferty, A.K. McCallum and F. Pereira, “Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data,” in *International Conference on Machine Learning*, 2001.
- [20] A. Esuli and F. Sebastiani, “SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining,” in *LREC*, 2006.
- [21] K. Krippendorff, “Estimating the reliability, systematic error, and random error of interval data,” *Educational and Psychological Measurement*, vol. 30 (1), pp. 61-70, 1970.
- [22] K. Krippendorff, *Content analysis: An introduction to its methodology*, Thousand Oaks, CA: Sage, 2004.
- [23] P. Ekman, “Facial expression and emotion,” *American Psychologist*, vol. 48(4), pp. 384-392, 1993.
- [24] B. Liu, “The challenge is still the accuracy of sentiment prediction and solving the associated problems,” in *5th Annual Text Analytics Summit*, 2009.