

Identifying Emotion Holder and Topic from Bengali Emotional Sentences

Dipankar Das

Department of Computer Science & Engineering
Jadavpur University
Kolkata – 700032, INDIA
dipankar.dipnil2005@gmail.com

Sivaji Bandyopadhyay

Department of Computer Science & Engineering
Jadavpur University
Kolkata – 700032, INDIA
sivaji_cse_ju@yahoo.com

Abstract

This paper presents two different approaches that are adopted for identifying emotion holders and topics from annotated Bengali blog sentences. The baseline system is developed based on the part-of-speech (*POS*), Name Entities (*NEs*) and noun phrases that are referred as the emotional expressions. An unsupervised syntactic system is developed based on the argument structure of the sentences with respect to its verb. If the acquired argument structure of a Bengali blog sentence with respect to its verb matches with any of the frame syntax retrieved for its equivalent English verb of identical sense from the VerbNet, the holder and topic related thematic roles associated with the English VerbNet frame are mapped to the appropriate phrases in the Bengali sentence. Simple rule based error reduction techniques are employed to improve the performance of the syntactic system significantly on the development sentences. The syntactic system with the average *F-scores* of 66.03% and 61.98% outperforms the baseline system with the average *F-scores* of 53.85% and 50.02% for single and multiple emotion holders and topics respectively on 500 gold standard test sentences.

1 Introduction

The source or holder of an emotional expression described in natural language text is the speaker or writer or experiencer (Wiebe *et al.*, 2005). Researches on emotion holder extraction are important for discriminating emotions that are viewed from different perspectives (Seki, 2007). By grouping opinion or 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 (Kim and Hovy, 2006).

On the other hand, topic is the real world object, event, or abstract entity that is the primary subject of the emotion or opinion as intended by the holder. Topic depends on the context in which its associated emotional expression occurs (Stoyanov and Cardie, 2008b).

Holder and topic described in natural language texts play the important roles with respect to the emotion of a reader or writer (Yang *et al.*, 2009). A wide range of natural language processing (*NLP*) tasks such as tracking users' emotion about reviews or events or politics as expressed in online forums or news, customer relationship management are using the emotional information.

Emails, weblogs, chat rooms, online forums and even twitter are considered as the affective communication substrates to analyze the reaction of emotional catalysts. Blog is one of the crucial communicative and informative repositories of text based emotional contents in the Web 2.0 (Lin *et al.*, 2007). Many blogs act as online diary of the bloggers reporting daily activities. Sometimes, blog posts are annotated by other bloggers. The blog is therefore considered for analyzing the reactions of users' emotions associated with various topics.

In the present task, the identification of emotion holder and topics is attempted for Bengali; a less privileged, less computerized and morphologically rich language. We have considered the Bengali blog corpus (Das and Bandyopadhyay, 2010c) that is annotated with emotion holders and topics along with Ekman's (1993) six basic emotional expressions and intensities at sentence level.

The blog sentences are passed through an open source Bengali shallow parser¹. The shallow parser gives different morphological information (root, lexical category of the root, gender, number, person, case, vibhakti, tam, suffixes etc.) that help in identifying the lexical patterns such as different part-of-speech (*POS*) combinations, Name Entities (NEs) and noun phrases. The lexical pattern based phrase level similarity clues are considered as the probable candidates for emotion holders and topics in the baseline system. But, the sentences containing implicit holders and or with passive occurrences trouble the baseline system. The baseline system achieves the average *F-scores* of 53.85% and 50.02% for identifying emotion holders and topics respectively on 500 test sentences.

Developing the syntactic model based on argument structures of the sentences satisfies the demand of the baseline system. The pivotal hypothesis considered in the syntactic model is based on the hypothesis followed in (Das *et al.*, 2009; Banerjee *et al.*, 2009). The verb based argument structures are acquired from the chunk or phrase level lexical patterns of the shallow parsed blog sentences. Equivalent English verbs of identical sense for the Bengali verbs are retrieved using Bengali to English bilingual dictionary². The available frames of the equivalent English verbs are extracted from English VerbNet (Kipper-Schuler, 2005). If any of the acquired argument structures matches with any of the extracted VerbNet frames of a sentence, the thematic roles for holder (e.g. *Experiencer*, *Agent*, *Actor*, *Beneficiary* etc.) and topic (e.g. *Topic*, *Theme*, *Event* etc.) associated with the frame are mapped to the appropriate slots in the Bengali sentence. But, the empirical evaluation on 600 development sentences shows that though the syntactic model outperforms the baseline system, the syntactic system suffers in case of resolving some errors (e.g. appositive cases, co reference with emotional expression, multiple holders and topics, overlapping topic spans, anaphoric presence of the holders). Thus, some simple rule based error reduction techniques based on rhetorical structure, emotional expressions of the sentences are added with the syntactic system. The performance of the syntactic system improves significantly on the development set. The average *F-scores* of the syntactic system are

66.03% and 61.98% for single as well as multiple emotion holders and topics respectively on gold standard 500 test sentences.

The rest of the paper is organized as follows. Section 2 describes the related work. The preprocessing of the corpus is described in Section 3. The baseline and syntactic systems are discussed in Section 4 and Section 5 respectively. Evaluation results along with associated errors and corresponding reduction techniques are mentioned in Section 6. Finally Section 7 concludes the paper.

2 Related Work

The survey on emotion holder and topic identification shows that the research has a wide scope except a few well-known works. Prior work on opinion holders has sometimes identified only a single opinion per sentence (Bethard *et al.*, 2004) and sometimes several (Choi *et al.*, 2005). Another work on labeling the arguments of the verbs with their semantic roles using a novel frame matching technique was carried out in (Swier and Stevenson, 2004). The Multi Perspective Question Answering (MPQA) with supporting annotation task was attempted in (Wiebe *et al.*, 2005). Based on the traditional perspectives, the task discussed in (Hu *et al.*, 2006) uses an emotion knowledge base for extracting the emotion holders. The machine learning based classification task for “not holder”, “weak holder”, “medium holder”, or “strong holder” are described in (Evans, 2007). An anaphor resolution based opinion holder identification method exploiting lexical and syntactic information from online news documents was attempted in (Kim *et al.*, 2007). The syntactic models of identifying emotion holder for English emotional verbs are developed in (Das and Bandyopadhyay, 2010a).

In the related area of opinion topic extraction, different researchers contributed their efforts. Some of the works are mentioned in (Kobayashi *et al.*, 2004; Popescu and Etzioni, 2005; Yi *et al.*, 2007). But, all these works are based on lexicon look up and are applied on the domain of product reviews. The topic annotation task on the MPQA corpus is described in (Stoyanov and Cardie, 2008a). The authors have pointed out that the target spans alone are insufficient for many applications as they neither contain information indicating which opinions are about the same topic, nor provide a concise textual representation of the topics.

¹ http://ltrc.iiit.ac.in/showfile.php?filename=downloads/shallow_parser.php

² <http://home.uchicago.edu/~cbs2/banglainstruction.html>

Kim and Hovy (2006) identified opinion holder with topic from media text using semantic role labeling framework. Another work on finding opinion sources and targets is discussed in (Ruppenhofer *et al.*, 2008). The above works are closely related to the present one.

All the above works have been attempted for English or Chinese. Recent study shows that non-native English speakers support the growing use of the Internet³ and the rapidly growing web users from multilingual communities focus the attention to improve the multilingual search engines on the basis of sentiment or emotion. To the best of our knowledge, Bengali is the sixth popular language in the world⁴, second in India and the national language of Bangladesh. At present, no work on emotion holder and or topic identification has been carried out for Bengali and even for any Indian language. Thus we believe that the present task and methodologies for identifying emotion holder and topic would help in the development of emotion analysis systems for other languages as well. The present approach aims to acquire all probable emotion holders and topics from a sentence if multiple occurrences are present in the sentence.

3 Pre-Processing of Corpus

A portion of the whole annotated blog corpus (Das and Bandyopadhyay, 2010c) containing 1100 sentences is considered in the present task. Each sentence of the corpus is annotated with emotional components such as emotional expression (word/phrase), intensity, associated holder and topic(s). Ekman's six emotion classes (*anger, disgust, fear, happy, sad* and *surprise*) along with three types of intensities (*high, general* and *low*) are considered for the sentence level annotation. As the sentences were collected from the blog posts, the *writers* of the blog posts are assumed as the default additional emotion holders (Wiebe *et al.*, 2005). If a sentence contains multiple emotional expressions, respective emotion holders and topics associated with each of the expressions are annotated in the corpus. It is observed that the long sentences contain more than one emotional expression and hence associated with multiple emotion holders and topics. All the probable emotion holders of a sentence are stored in an anchoring vector of the annotated corpus. If multiple emotion holders exist, the

successive holders are annotated and placed in an anchoring vector according to their order of occurrence in the respective sentence. But the annotated topic spans are simply mentioned by specifying the text segments. The following Bengali sentence in Example 1 shows the emotional expression, its associated holder and topic.

Example 1:

রাশেদ বলেছেন আপনার কবিতাটা
(*Rashed*) (*bolechen*) (*apnar*) (*kabitata*)
পড়তে গিয়ে তার এই সুন্দর কৌতুকটা
(*porte*) (*giye*) (*tar*) (*ei*) (*sundar*) (*koutukta*)
মনে পড়ছিলো।
(*mone*) (*porchilo*).

Rashed said that he was remembering this **beautiful comic** while reading your **poem**.

Emotional Expression: সুন্দর কৌতুক *sundar koutuk* '**beautiful comic**' Holder: < *writer*, রাশেদ '*Rashed*' > Topic: কবিতা *kabita* '**poem**'

The emotional blog sentences are passed through an open source Bengali shallow parser⁵. The shallow parser gives different morphological information (root, lexical category of the root, gender, number, person, case, vibhakti, tam, suffixes etc.) that help in identifying the lexical as well as phrase level syntactic patterns of the sentences. The lexical phrase level pattern of the shallow parsed emotional expression of Example 1 is shown in Figure 1.

(((JJP সুন্দর JJ <fs af='সুন্দর ,adj,,,,d,শুন্য,শুন্য'>)
(NP কৌতুকটা NN <fs af='কৌতুক ,v,,,,, টা , টা >))

Figure 1: Shallow parsed phrasal pattern of an emotional expression

Both the baseline and syntactic systems use the shallow parsed sentences to identify the holders and topics associated with the corresponding emotional expressions. The shallow parsed sentences are pre-processed to generate the simplified patterns. The word level POS patterns and chunk level phrasal patterns are helpful for the baseline and syntactic systems respectively.

³ <http://www.internetworldstats.com/stats.htm>

⁴ http://www.ethnologue.com/ethno_docs/distribution.asp?by=size

⁵ http://ltrc.iit.ac.in/showfile.php?filename=downloads/shallow_parser.php

4 Baseline System

The baseline system is developed based on the lexical level similar patterns extracted from the simplified shallow parsed sentences. The similar patterns are grouped according to the part-of-speech (POS) categories. It is observed that the hints are present mostly in the sentences containing comments of the blog users. Each of the sentences containing users' comments is started with a corresponding username. The username is the default hint that helps in capturing the first holder present in an anchoring vector of nested sources.

In other cases, the POS tagged sentences contain the similarity pattern at lexical level. The words that are tagged as Named Entities (NEs), commonly in the form of NNPC (Compound proper noun), NNP (Proper noun), NNC (Compound common noun), NN (Common noun) or PRP (Pronoun) present at the beginning of a sentence are considered as the responsible candidates for emotion holder. The similarity pattern consists of two phrasal constituents, the subject and the verb. The rest common portions (*Common_Portion*) containing the additional constituents are basically the floating portions. As Bengali is a free phrase order language, the ordering between the verb and the floating portion is not fixed.

The general POS level pattern such as [*<NNP /NNPC /NN /NNC /PRP > {<VBZ /VM> <Common_Portion>*] is considered for capturing the clue of an emotion holder. The components of the common portion are assembled starting with the hint of the first occurring POS tags of types NNP or NNC or PRP in the POS tagged sentence. Reaching at the verb based POS tags like VBZ or VM, the system stops incorporating the components into the common portion. The rest of any component present in the sentence after the verb is therefore simply appended to finally build the common portion. If no emotion holder is identified from the sentences, the default emotion holder, i.e. *writer* is chosen as an emotion holder.

A preliminary investigation on 600 development sentences reveals that all the identified emotion holders are not directly linked with the emotional expressions for which they are annotated in the gold standard sentences. To avoid such misleading strategy, we have additionally incorporated the knowledge of emotional expressions for identifying the emotion holders. We extract all component words from the chunks

that contain at least one emotion word (e.g. কৌতুক *koutuk* 'comic'), and match them against the Bengali *WordNet Affect* lists (Das and Bandyopadhyay, 2010b). The words present in the extracted chunks are then treated as the candidates for emotional expressions. Identification of an emotional expression containing a single emotion word is straightforward. Consecutive words that appear in the chunks and contain at least one emotion word are identified as the long emotional expressions. The chunks that occur as the immediate neighbors of the emotional expressions are tagged as the responsible candidates for emotion holders.

The shallow chunked phrases that are formed by removing emotional expressions and holders are identified as the responsible spans containing one or more potential emotion topics. The words of the shallow chunks containing POS tags of NNP or NNC or noun are also possible emotion topics. The *Common_Portion* has been considered as the text span for containing topics. The chunks that occur as the immediate neighbors of the verb chunks or emotional expressions are tagged as emotion topic spans. As the holders and topic spans are identified at chunk or phrase level, the tagged entities are matched against full word strings or phrases. There is no question of partial match or head match (Lu, 2010).

It is observed that the similarity patterns exist mostly in the simple active sentences that contain single emotion holder and topic. As the identification from long complex or compound sentences shows problems in the baseline system, the system fails to capture the nested emotion holders or multiple topics. The topics containing single word tokens have been identified easily rather than multiword topic spans. We therefore turn our focus towards developing syntactic system to handle the active sentences as well as multiple holders and potential topics. The baseline system achieves low average *F-scores* of 53.85% and 50.02% for emotion holder and topic respectively on 500 gold standard test sentences.

5 Syntactic System

The syntactic way of identifying argument structure of a sentence and capturing the emotion holders and topics from the viewpoint of *thematic roles* has therefore been considered as a favored way to meet up the demands of the baseline model. More specifically, the argument structure or subcategorization frame of a sen-

tence with respect to its verb plays a crucial role in identifying the holders and topics.

A subcategorization frame is a statement of what types of syntactic arguments a verb (or an adjective) takes, such as objects, infinitives, that-clauses, participial clauses, and subcategorized prepositional phrases (Manning, 1993). The hypothesis that was considered in (Banerjee *et al.*, 2009) for extracting Bengali subcategorization frames is also considered in the present task. The hypothesis is that the verb subcategorization frames or argument structure for the equivalent English verbs (sharing the same sense) of a Bengali verb are the initial set of valid verb subcategorization frames for that Bengali verb. Irrespective of ordering, the phrase or chunk level similarities between Bengali and English languages help in acquiring the equivalent argument structures of the Bengali sentences. The argument structures of the Bengali emotional sentences contain phrase level head information that in turn conveys the sentential syntax information.

5.1 Bengali Verb Identification

As no full-fledged parser is available in Bengali, the shallow parsed sentences containing phrase level chunks are used for identifying the emotion holders and topics.

To identify the simple verbs from the chunked corpus, the words that are tagged as main verb (VM) and belong to the verb group chunk (VGNF) in the corpus are identified (e.g. ভালোবাসা *bhalobasa* ‘love’). For compound or conjunct verbs, the pattern such as {[XXX] (NN) [YYY] (VM)} are retrieved from the Bengali chunked corpus (e.g. VGNF {[আনন্দ *ananda*] (NN) [করা *kara*] (VM)} means *enjoy*). The light verb [YYY] tagged with ‘VM’ generally occurs in any inflected form. Different suffixes may be attached to a simple verb or light verb depending on various features like Tense, Aspect, and Person. Hence, the passive sentences are also handled by incorporating the verb specific knowledge extracted from shallow parsed result. A Bengali stemmer with an accuracy of 97.09% uses a suffix list to identify the stem form of the Bengali simple verbs and light verbs. Another knowledge base stores the stem forms and the corresponding dictionary forms of 374 Bengali verbs containing simple and light verb entries. The dictionary forms of the Bengali compound or conjunct verbs are made by incorporating the dictionary forms of the light verbs with their pre-

ceding noun words that are tagged as ‘NN’ present in the retrieved lexical patterns.

5.2 English Verb Determination

The determination of equivalent English verbs of an identified Bengali verb is carried out using a Bengali to English bilingual dictionary⁶. The method to extract the English equivalent synsets of the Bengali verbs is based on the work done in (Banerjee *et al.*, 2009). We have identified the equivalent English verb synsets of the Bengali verb entries that are present in the bilingual dictionary. For example, the dictionary entries for simple verb ভালোবাসা *bhalobasa* ‘love’ and conjunct verb আনন্দ করা *ananda kara* ‘enjoy’ are as follows.

< ভালোবাসা [bhālōbāsā] v to *love*, to be amorous to wards; to *like*; to have attachment or affection, fondness for ...>

< আনন্দ করা v. to *rejoice*; to *make merry*....>

Different synonyms for a Bengali verb having the same sense are separated using “,” and different senses are separated using “;” in the dictionary. The synonyms including similar senses of the target verb are extracted from the dictionary and yield a resulting set called English Equivalent Synset (EES). For example, two English Equivalent Synsets (EES) are extracted for the conjunct verb আনন্দ করা *ananda kara* ‘enjoy’.

5.3 Retrieval of English VerbNet Frames

It is found that each of the English Equivalent Synsets (EES) occur in each separate class of English VerbNet (Kipper-Schuler, 2005). VerbNet associates the semantics of a verb with its syntactic frames and combines traditional lexical semantic information such as *thematic roles* and *semantic predicates* with *selectional restrictions*. Member verbs in the same VerbNet class share common syntactic frames and thus they are believed to have the same syntactic behavior. The VerbNet files containing member verbs and possible subcategorization frames are stored in XML file format. Hence, the XML files of VerbNet are pre-processed to build up a general list that contains all verbs, their classes and possible subcategorization frames (primary as well as secondary).

⁶ <http://home.uchicago.edu/~cbs2/banglainstruction.html>

This pre-processed list is searched to retrieve the present subcategorization frames for each verb (e.g. *love*) of the English Equivalent Synsets (EES) (e.g. *love*) corresponding to the Bengali verb. These retrieved subcategorization frames are believed to be the valid set of argument structures for the Bengali verbs (Das *et al.* 2009; Banerjee *et al.*, 2010). The retrieved subcategorization frames for each sentence are now used for mapping with the acquired Bengali argument structure. The acquisition process is discussed in Section 5.4.

5.4 Mapping of Acquired Bengali Argument Structures

The shallow chunked Bengali sentences are passed through a rule based *phrasal-head* extraction module to identify the phrase level argument structures of the sentences corresponding to the position of the verbs. The extracted *head part* of every phrase from the chunked sentence is considered as a component of the argument structure of that sentence.

If the acquired argument structure for a Bengali emotional sentence is matched with any of the available extracted frames of English VerbNet, the thematic information with respect to emotion holder (e.g. *Experiencer*, *Agent*, *Actor*, etc.) and topic (e.g. *Topic*, *Theme*, *Event* etc.) associated with the English frame syntax are mapped to the appropriate slots of the acquired Bengali argument structure. Tag conversion routines are developed to transform the POS of the system-generated argument structures into the POS of the VerbNet frames.

For example, in simple sentences (as shown in Example 2), the occurrence of the NNPC, NNP, NNC or NN tags preceded by the PRP (Pronoun) NNP, NNC, NN or NNPC tags (may contain case markers (e.g. কে [-ke]) and followed by a verb gives similar frame syntax as of the “Basic Transitive” frame of English VerbNet.

The acquired argument structure of another Bengali emotional sentence in Example 3 shows that the argument structure contains sentential complement “S” started by যে -je with DET type POS. The argument structure is acquired for the Bengali verb অনুভব করা *anubhab kara* ‘feel’. One of the extracted VerbNet frame syntax containing -that type sentential complement for the equivalent English verb *feel* is shown in Table 1. As the acquired argument structure matches with the extracted VerbNet frame syntax, the emotion holder related thematic role (e.g. *Experiencer*)

associated with the VerbNet frames are mapped to the equivalent phrases (রাশেদ *Rashed*) of the acquired argument structure of the Bengali sentence.

Example 2:

Holder: < writer, রাম ‘Ram’>

Topic:< সিতা ‘Sita’>

রাম সিতাকে ভালবাসে
(Ram) (Sitake) (bhalobase)
Ram loves Sita.

Acquired Argument Structure: [NNP NNP-ke VM]

Extracted VerbNet Frame Syntax: [<NP value=“Experiencer” ></VERB><NP-theme>]

Example 3:

Holder: < writer, রাশেদ, ‘Rashed’, রাম ‘Ram’>

Topic:< সুখ ‘sukh’ >

রাশেদ অনুভব করেছিল যে
(*Rashed*) (anubhab) (korechilo) (je)
রামের সুখ অনন্তহীন
(Ramer) (sukh) (antohin)
Rashed felt that Ram’s pleasure is endless.

Acquired Argument Structure: [NNP VM DET-je S]

Extracted VerbNet Frame Syntax: [<NP value=“Experiencer” ></VERB>< S-that (Sentential -that Complement)>]

Example 4:

Holder: < writer, সায়ণ ‘Sayan’>

Topic:< -- >

সায়ণ ভীষণ আনন্দ অনুভব
(Sayan) (bhishon) (anondo) (anubhab)
করেছিল
(korechilo)
Sayan felt very happy.

Acquired Argument Structure: [NNP NN VM]

Extracted VerbNet Frame Syntax: [<NP value=“Experiencer” ></VERB>< NP-theme>]

Table 1: Example sentences, holders, topics, acquired argument structures and extracted VerbNet frames

As the sentence contains a sentential complement as its arguments, the portion of sentential complement is passed by the syntactic system to obtain any implicit emotion holder or topic if present. The sentential complement contains the topic (সুখ *sukh* ‘pleasure’). The case markers in Bengali are required to identify the emotion holders as the case markers give the useful hints to capture the *selectional restrictions* that play a key role in distinguishing the emotion holders from other valid alternatives.

6 Evaluation

The evaluation results of the baseline and syntax based systems for single and multiple emotion holders and topics are presented in Table 3. The baseline system performs satisfactorily for the simple sentences containing single emotion holder. The error analysis suggests that the rich morphology and free phrase order nature of Bengali restricts the baseline system to capture the holder information. The baseline system suffers in disambiguating the emotional holders for complex and compound sentences, as no full-fledged dependency parser is available in Bengali.

Hence, we explore the syntactic system based on the shallow chunked results. The holder and topic identification based on argument structure improves the syntactic system as it compares the phrasal similarity patterns from phrasal heads rather than only POS tag combinations.

Argument structure acquisition from chunked sentences contributes effectively to holder and topic identification task. The result shows that the syntactic system outperforms the baseline system reasonably. Distinguishing arguments from adjuncts and the holder identification task for passive sentences are not handled in the baseline system whereas the case markers and verb specific information are used by the syntactic system to handle such typical instances.

The error analysis has been conducted for the syntactic system on 600 development sentences. It is observed that the system fails to capture the following cases, mostly to resolve the co reference problems in blog texts.

Case 1. Appositive Use: The implicit emotion holders present in a sentence. (e.g. রাম ‘Ram’ in case of রামের সুখ ‘Ram’s pleasure’). The identification of emotion holder at sentence level requires the knowledge of two basic constraints (*implicit* and *explicit*) separately. The *explicit* constraints identify single prominent *emotion*

holder that is directly involved with the emotional expression whereas the *implicit* constraints identify all direct and indirect nested sources as *emotion holders*. The following example contains the emotion holder নাসরিন সুলতানা (*Nasreen Sultana*) based on *implicit* constraints.

Example 5:

Holder: <গেদু চাচা, নাসরিন সুলতানা >

গেদু চাচা বলে, না গো বোন, আমি
(Gedu ChaCha) (bole) : (na) (go) (bon), (ami)

নাসরিন সুলতানার দুঃখের কথাতে
(Nasreen Sultana) (dookher) (kathate)

কেঁদে ফেলি।

(kende) (feli)

Gedu Chacha says, no my sister, I fall into cry on the sad speech of Nasreen Sultana

Solution: We have considered the suffixes that are determined from the shallow parsed phrases to identify the appositive cases. In Example 3, the appositive case (e.g. রামের সুখ (*Ram’s pleasure*)) is also identified and placed in the vector by removing the inflectional suffix (-এর *-er* in this case). Sometimes, the vibhakti and tam information also play the effective roles in identifying emotion holders.

Case 2. Co-reference with Emotional Expression: It is observed that the system captures the emotion holders and topics but sometimes, the identified holders and topics are not referred by the emotional expressions. More specifically, they do not contain any direct linking with emotion. In the following example, রাশেদ *Rashed* is only the emotion holder of the emotional expression রাগ *rag* ‘angry’. But, system additionally identifies রিতু *Ritu* that is not an emotion holder in the present case.

Example 6:

রিতু যা কথা বলে, রাশেদ তাতেই

(Ritu) (ja) (katha) (bole), (Rashed) (tatei)

রেগে যায়।

(rege) (jai)

Rashed gets angry on whatever Ritu says.

Solution: It is necessary to identify the emotional expressions that are referred by the holders and topics. To accomplish the goal, we extract all component words from the shallow chunks that contain at least one emotion word (e.g. কৌতুক *koutuk* ‘comic’), and match them against the

Bengali *WordNet Affect* lists (Das and Bandyopadhyay, 2010b). The identification of emotional expressions containing single word as well as long multi words are carried out in a similar way that is adopted in the baseline system of Section 4. The immediate neighboring chunks of the identified emotional expressions are only considered as co referred chunks. The co referred chunks responsible for holder and topic are now tagged by the syntactic system.

Case 3. Multiple Holders and Topics: The complex or compound sentences contain more than one clause and each of the clauses may contain individual emotional expression. The holders and topics associated with the emotional expressions in all the clauses need the fine-grained study of the sentential structures rather than depending on simple subcategorization knowledge. The following example shows that two emotional expressions (দুঃখ *dookkha* ‘sorrow’ and আনন্দ *ananda* ‘happy’) contain two different emotion holders (গেদু চাচা *Gedu ChaCha* and চাচি *Chachi*).

Example 7:

গেদু চাচার দুঃখ থাকা সত্ত্বেও
(Gedu ChaChar) (dookkha) (thaka) (satweo)
চাচি আনন্দ করে সবাইকে
(Chachi) (ananda) (kare) (sabaikē)
নিয়ে থাকে ।
(niye) (thake)

Though Gedu Chacha has sorrow, Chachi lives happily with all.

Solution: As the complex or compound sentences contain more than one clause and each of the cluses contains individual emotional expressions, we consider the sentential rhetorical structure. Instead of identifying rhetorical relations (Mann and Thompson, 1988), the present task acquires the rhetorical components such as *locus*, *nucleus* and *satellite* from a sentence as these rhetoric clues help in identifying the individual holder and topics associated in each clause of the sentences.

The part of the text span containing the emotional expression is considered as *locus*. Primarily, the separation of *nucleus* from *satellite* is done based on the punctuation markers (,) (!) (?). Frequently used *discourse markers* (যেহেতু *jehetu* ‘as’, যেমন *jemon* ‘e.g.’, কারণ *karon* ‘because’, মানে *mane* ‘means’) and *causal verbs* (ঘটায় *ghotay* ‘caused’) are also the useful clues

if they are explicitly specified in the text and present in a manually prepared seed list. If any word in the emotional expression co-occurs with any word element of the *nucleus* or *satellite* in the same chunk, the feature is considered as *common rhetoric similarity*. Otherwise, the feature is considered as *distinctive rhetoric similarity*. The chunks identified by the syntactic system as holder and topic and tagged as *common rhetoric similarity* are only considered for each of the clauses of a sentence. For the very reason, all possible holders and topics associated in all the clauses of a sentence are identified by the syntactic system.

Case 4. Overlapping Topic Spans: It is observed that the emotion topics containing single word tokens have been identified easily rather than multi word topics. Sometimes, the emotion related topics co-exist with other potential non-emotional topics. As the topics may consist of multi word strings, the text spans denoting the topic spans create problems in identifying emotion topic span from other non-emotional topic spans. In the following example, the emotional expression আনন্দ *ananda* ‘enjoy’ is related with the topic গান *gan* ‘song’ and টিভি *TV* ‘television’. The syntactic system additionally captures বই *boi* ‘book’ that is a potential but non emotion topic.

Example 8:

তুমি তো বই পড়তেই না, এখন
(tumi) (to) (boi) (portei) (na), (ekhon)
দেখছি তুমি গান, টিভি তেও
(dekhchi) (tumi) (gan), (TV) (teo)
আনন্দ পাওনা।
(ananda) (paona)

You never used to read books, now we notice that you do not enjoy in song or television too.

The topic of an opinion depends on the context in which its associated opinion expression occurs (Stoyanov and Cardie, 2008a). The *common rhetoric similarity* feature helps the syntactic system by aiming to separate emotion topics from non-emotion topics as well as to separate the overlapping possibilities of discrete emotion topic spans from non-topical contiguous regions. If the identified topic chunks are tagged with *common rhetoric similarity*, the chunks are classified as emotional topics and separated from non-topical elements in a sentence.

Cases	Syntactic System			
	Single Holder	Multiple Holder	Single Topic	Multiple Topic
Before Error Analysis	57.83	53.76	54.62	53.96
Case 1	60.78	56.11	54.62	53.96
Case 1+Case 2	62.44	59.12	57.90	55.25
Case 1+Case 2+Case 3	66.82	62.53	61.48	57.96
Case 1+Case 2+ Case 3+Case 4	66.82	62.53	65.77	60.58
Case 1+Case 2+ Case 3+Case 4+Case 5	69.16	66.24	65.77	60.58

Table 2. *F-scores* (in %) of the different error cases of syntactic system on development set

Case 5. Anaphoric Presence of Holders: The emotion holders are sometimes referred using anaphors. Sometimes, the candidate anaphors are linked with the emotional expressions instead of the actual emotion holders. The problem is identified in the Example 5 shown earlier. The actual emotion holder গেদু চাচা ‘Gedu ChaCha’ expresses the emotion in a clause that is represented by the anaphor আমি ami ‘I’ in another clause.

Solution: The sentences of user comments in the adopted blog corpus contain a special default phrasal pattern that helps in identifying the emotion holders ([<Named Entity> <say>] e.g. গেদু চাচা বলে: (*Gedu ChaCha bole*), রাশেদ বলেছেন (*Rashed bolechen*), সায়ন বলেছে (*Sayan bolechhe*) etc.). Hence, if a pronoun is present with an emotional expression, the preceding Named Entities of such default phrasal pattern have been considered as the emotion holders.

Incorporating different simple rule based techniques has reduced the errors of the above problems. The inclusion of such error removing techniques has improved the performance of the syntactic system in terms of *F-score*. The result for each of the cases is shown in Table 2. As it is observed from the evaluation on the development sentences that the application of the individual error reduction techniques does not affect the role of other, we have added the error reduction techniques one after another and finally achieved the best results in terms of *F-score*.

The comparative results of the baseline and syntactic systems for single as well as multiple emotion holders and topics on 500 test sentences are shown in Table 3. It has been observed that though the syntactic system outperforms the baseline, both the systems perform well in case of identifying single holder and topic rather than

multiple holders and topics. The reason may be the metaphoric use of emotion or the ungrammatical structure of some blog sentences.

Categories	Baseline		Syntactic	
	Single	Multiple	Single	Multiple
Holder	55.67	52.17	68.03	64.21
Topic	52.97	49.01	64.04	59.34

Table 3. *F-scores* (in %) of the Baseline and Syntactic systems on the test set

7 Conclusion

The paper describes the emotion holder and topic identification in Bengali based on the argument structure. The evaluation result suggests for incorporating dependency parsing to improve the *F-score* along with the handling of inflections to extract more frames. The adoption of supervised techniques is the future task to accomplish. The further exploration of holder identification task is aimed for tracking emotion from the perspectives of single as well as multiple users on basis of time and topic. The identification of metaphoric usages and discourse in the natural language texts is the future area to be explored from the perspective of emotion.

References

- Banerjee S., D.Das and S.Bandyopadhyay. 2010. Classification of Verbs – Towards Developing a Bengali Verb Subcategorization Lexicon. *Global WordNet Conference-2010*, pp. 76-83, Mumbai, India.
- Banerjee S., D.Das and S.Bandyopadhyay. 2009. Bengali Verb Subcategorization Frame Acquisition - A Baseline Model. *Workshop on Asian Language Resource-7, Association of Computational Linguistics*

- tics (ACL) and International Joint Conference on Natural Language Processing (IJCNLP), pp. 76-83, Suntec, Singapore.
- Bethard Steven., H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky. 2004. Automatic Extraction of Opinion Propositions and their Holders. In *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*.
- Choi, Y., C. Cardie, E. Riloff, and S. Patwardhan. 2005. Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. In *Proceedings of HLT/EMNLP*.
- Das D., A. Ekbal, and S. Bandyopadhyay. 2009. Acquiring Verb Subcategorization Frames in Bengali from Corpora. In *the proceedings of the 22nd International Conference on the Computer Processing of Oriental Languages (ICCPOL-09)*, LNAI-5459, pp.386-393, Hong Kong.
- Das D. and S. Bandyopadhyay. 2010a. Emotion Holder for Emotional Verbs – The role of Subject and Syntax. *CICLing- 2010*, A. Gelbukh (Ed.), LNCS 6008, pp. 385–393, Romania.
- Das D. and S. Bandyopadhyay. 2010b. Developing Bengali WordNet Affect for Analyzing Emotion. In *the proceedings of the 23rd International Conference on the Computer Processing of Oriental Languages (ICCPOL-2010)*, pp. 35-40, California, USA.
- Das D. and S. Bandyopadhyay. 2010c. Labeling Emotion in Bengali Blog Corpus – A Fine Grained Tagging at Sentence Level. In the 8th Workshop on Asian Language Resources (ALR8), COLING-2010, pp. 47-55, August 21-22, Beijing, China.
- Ekman, P. 1993. Facial expression and emotion. *American Psychologist*, 48(4), pp.384–392
- Evans, D.K. 2007. A low-resources approach to Opinion Analysis: Machine Learning and Simple Approaches. *NTCIR*
- Hu, J., C. Guan, M. Wang, and F. Lin. 2006. Model of Emotional Holder. In *Shi, Z.-Z., Sadananda, R. (eds.) PRIMA 2006. LNCS (LNAI)*, vol. 4088, pp. 534–539. Springer, Heidelberg
- 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., E. Hovy. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. *ACL*.
- Kim, Y., Jung, Y., Myaeng, S.-H. 2007. Identifying Opinion Holders in Opinion Text from Online Newspapers. In *2007 IEEE International Conference on Granular Computing*, pp. 699–702, doi:10.1109/GrC.2007.45
- Kobayashi, N., K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima. 2004. Collecting evaluative expressions for opinion extraction. *IJCNLP*.
- Lin K. H.-Y., C. Yang and H.-H. Chen. 2007. What Emotions News Articles Trigger in Their Readers?. *SIGIR*, pp. 733-734.
- Lu Bin, 2010. Identifying Opinion Holders and Targets with Dependency Parser in Chinese News Texts. *Student Research Workshop, NAACL-HLT*, pp. 46-51, California.
- Mann, W. C. and S. Thompson. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization, *TEXT* vol. 8, pp.243–281
- Manning, C.D.1993. Automatic Acquisition of a Large Subcategorization Dictionary from Corpora. In *31st Meeting of the ACL*, Columbus, Ohio, pp. 235–242.
- Popescu, A. and O. Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of HLT/EMNLP*
- Ruppenhofer Josef, Swapna Somasundaran, and Janyce Wiebe. 2008. Finding the Sources and Targets of Subjective Expressions. *LREC 2008*.
- Seki, Y.2007. Opinion Holder Extraction from Author and Authority Viewpoints. In *SIGIR 2007. ACM*, New York, 978-1-59593-597-7/07/0007.
- Stoyanov, V. and C. Cardie. 2008a. Annotating topics of opinions, *LREC*.
- Stoyanov V. and C. Cardie. 2008b. Topic Identification for Fine-Grained Opinion Analysis, *Coling 2008*, pp. 817–824.
- Swier, R.S., S. Stevenson. 2004. Unsupervised Semantic Role Labelling. In *Proceedings of EMNLP*.
- Wiebe, J., T. Wilson, and C. Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation 1*(2).
- Yang Changhua, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. 2009. Writer Meets Reader: Emotion Analysis of Social Media from both the Writer's and Reader's Perspectives. *IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology Workshops*, pp. 286-290.
- Yi, J., T. Nasukawa, R. Bunescu, and W. Niblack. 2003. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of ICDM*.

Human-like holder plays an important role in identifying actual emotion expressed in text. This paper presents a baseline followed by syntactic approach for capturing emotion holders in the emotional sentences. The emotional verbs collected from WordNet Affect List (WAL) have been used in extracting the holder annotated emotional sentences from VerbNet. The baseline model is developed based on the subject information of the dependency-parsed emotional sentences. The unsupervised syntax based model is based on the relationship of the emotional verbs with their argument structure extracted from ...[^] Emotion Holder for Emotional Verbs - The Role of Subject and Syntax. 8 years 1 months ago.