Identifying Emotion Topic – An Unsupervised Hybrid Approach with Rhetorical Structure and Heuristic Classifier

Dipankar DAS
Department of Computer Science and Engineering, Jadavpur University
Kolkata, West Bengal, India
dipankar.dipnil2005@gmail.com

Sivaji BANDYOPADHYAY
Department of Computer Science and Engineering, Jadavpur University
Kolkata, West Bengal, India
sivaji_cse_ju@yahoo.com

Abstract:

This paper describes an unsupervised hybrid approach to identify emotion topic(s) from English blog sentences. The baseline system is based on *object* related dependency relations from parsed constituents. However, the inclusion of the *topic* related *thematic roles* present in the verb based syntactic argument structure improves the performance of the baseline system. The argument structures are extracted using VerbNet. The unsupervised hybrid approach consists of two phases; firstly, the information of *Rhetorical Structure* (RS) is extracted to identify the target span corresponding to the emotional expression from each sentence. Secondly, as an individual target span contains one or more topics corresponding to an emotional expression, a *Heuristic Classifier* (HC) is designed to identify each of the topic spans associated in the target span. The classifier uses the information of *Emotion Holder* (EH), *Named Entities* (NE) and four types of *Similarity* features to identify the phrase level components of the topic spans. The system achieves average *recall*, *precision* and *F-score* of 60.37%, 57.49% and 58.88% respectively with respect to all emotion classes on 500 annotated sentences containing single or multiple emotion topics.

Keywords:

Emotion Topic; Target Span; Topic Span; Rhetorical Structure; Heuristic Classifier; Similarity Feature

1. Introduction

Topic is the real world object, event, or abstract entity that is the primary subject of the opinion as intended by the opinion holder [20]. Topic span [20] associated with an opinion expression is the closest minimal span of text that mentions the topic and target span [20] is the text span that covers the syntactic surface form comprising the contents of the opinion. In our present task, the same definition and terminology are used for identifying topic span and target span. Two examples are given as follows.

Example 1.

"*He first cried up the* ***toy car***".

Example 2.

"*Max ignored the* ***issues of sports*** *as well as* ***politics***".

In Example 1, the sentence contains the topic span "*toy car*" with respect to the emotion expression "*cried up*" and the emotion holder "*he*". In Example 2, the sentence contains multiple topics as shown in bold face associated with the underlined target span. The topics are related to the emotional expression "*ignore*". The identification of topic span is difficult within the single target span of the opinion as there are multiple potential topics, each identified with its own topic span ([20], [23], [25]).

Although the identification of topic spans is difficult, the information of emotion topics is useful for the domain of Question Answering (QA), Information Retrieval (IR), product reviews, social media, stock markets, and customer relationship management. Major studies on Opinion Mining and Sentiment Analyses have been attempted with more focused perspectives rather than fine-grained emotions. Especially, the blog posts contain instant views, updated views or influenced views regarding single or multiple topics. Thus the present task deals with the identification of emotion topics from blog sentences.

In the present task, topics related to the emotional expressions are identified from the sentences of a blog corpus [19]. Three annotators have annotated the topic spans. The agreement of the annotated topic spans is measured using Cohen's kappa (κ) [28] and measure of agreement on set-valued items (MASI) [29]. The average agreements of 0.65 and 0.63 with respect to all emotion classes are obtained for kappa and MASI. Both the results show low agreements due to the problem of selecting single topic among multiple relevant alternatives as well as of identifying the scope of individual topic span.

The baseline system is developed based on the parsed constituents in the *object* related dependency relations. The system achieves an average *F-score* of 41.45% with respect to all emotion classes. The additional clues from *Topic* related *Thematic Roles* improve the performance of the baseline system. The phrase segments containing *Topic* as the *Thematic Role* are extracted from the verb

978-1-4244-6899-7/10/$26.00 ©2010 IEEE

based syntactical argument structures of the sentences. The argument structures are acquired from VerbNet [10]. But, the error analysis shows that the argument structures fail to capture the topic spans if multiple potential topics are present in the blog texts. In addition to that, the baseline system also suffers from identifying the boundary of each emotion topic.

Hence, a hybrid approach is adopted to identify multiple emotion topics along with their topic spans from each of the blog sentences. The approach consists of two phases, *Rhetorical Structure* (RS) extraction task followed by a *Heuristic Classification* (HC) technique. Rhetorical Structure Theory (RST) describes the various parts of a text, how they can be arranged and connected to form a whole text [12]. The theory maintains that consecutive discourse elements, termed *text spans*, which can be in the form of clauses, sentences, or units larger than sentences, are related by a relatively small set (20–25) of *rhetorical relations* ([13], [14]).

Topics are generally distributed in different text spans of writer's text and can be distinguished by capturing the rhetorical structure of the text. The writer's direct as well as indirect emotional intentions are reflected in the target span in a sentence by mentioning one or more emotional topics that point to the emotional expressions (word/phrase). Hence, the primary focus is to identify the target span in the sentence. RST distinguishes between the part of a text that realizes the primary goal of the writer, termed as *nucleus*, and the part that provides supplementary material, termed as *satellite.* Each sentence contains a *locus,* the main effective part of the *nucleus* or *satellite.*

The assumption in the present task is that the emotional expression is to be treated as the *locus.* T*he* components in the *nucleus* or *satellite* except the *locus* are used to identify the target span. The identification of the *locus* inside *nucleus* or *satellite* is carried out based on the *affect word* that is searched using the *WordNet Affect* [2] list. The identification of target span as well as the separation of *nucleus* segment from *satellite* is also carried out based on *direct* and *transitive* dependencies, *causal verbs*, *relaters* or *discourse markers* and emotion *holder.* Various rules are designed to identify the target span from the *nucleus* as well as from the *satellite.*

In Example 1, the emotional expression (*cried up*) and target span (*the toy car*) (in this case it is the also the topic span) are both present in a single clause and the target span can be identified by removing the *locus* (*cried up*) and emotion holder (*he*) from the sentence. But, there are some cases where the *nucleus-satellite* combination is used to identify the target span and to determine whether or not a rhetorical structure holds between two spans of text. In Example 3, the sentence contains two potential topics "*summer vacation*" and "*play cricket*" in the *nucleus* and *satellite* respectively. Hence, the target span consists of both *nucleus* and *satellite*. The portions of *nucleus* and *satellite* are marked as "{}" and "[ ]" and the topics are shown in bold face in the examples.

Example 3.

"{I *enjoyed* the **summer vacation**} [because I had a golden chance to **play cricket** in that period].

Example 4.

"{I am currently *angry*}[because I want **Jarred** to take me for **DQ Blizzard**].

The sentences are passed through a chunker [26] to identify the phrase level components that help in identifying individual topic spans inside a target span. An unsupervised *Heuristic Classifier* (HC) is applied on the chunked target span to identify one or more topic spans. Each of the topic spans is assigned with a *heuristic score* that is calculated from the same *feature scores.* The features used in the classification are *Emotion Holder (EH)*, *Named Entities* (NE), *Structural Similarity, Sentiment Similarity, Syntactic Similarity* and *Semantic Similarity.* For example, *Semantic Similarity* contains different features (e.g. *synonym*, *hypernym*, *SenseID*) of WordNet [7] with respect to the *locus*. The phrases with *heuristic scores* (> *0.5*) are selected as the potential topic spans. The *recall* (60.37%) of the system is better than the *precision* (57.49%) as the system is strong enough to identify one or more potential emotion topics from the sentences but suffers in determining the scopes or spans of the topics to some extent.

The rest of the paper is organized as follows. Section 2 describes the related work done in this area. The topic annotated corpus preparation is discussed in Section 3. The baseline system is described in Section 4. The hybrid technique containing *Rhetorical Structure Extraction* (RSE) and *Heuristic Classification* (HC) is discussed in Section 5. Evaluation results are specified in Section 6. Finally Section 7 concludes the paper.

2. Related Work

The method of identifying an opinion with its holder and topic from online news is described in [9]. The model extracts opinion topics for subjective expressions signaled by verbs and adjectives. They have extracted the topics associated with a specific argument position from the verb or adjective based argument structures. Similarly, the verb based argument extraction and associated topic identification is followed in the present baseline system. As the baseline system suffers in identifying argument structures from blog texts, the incorporation of the knowledge regarding rhetorical structure improves the performance.

In the related area of opinion topic extraction, different researchers contributed their efforts. Some of the works are mentioned in ([27], [11], [17]). 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 ([20], [24, [25]).

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. The present task aims to capture the topic related inherent structure from blog texts by the application of rhetorical structure knowledge. The contribution of rhetorical structure knowledge helps in identifying more focused target span associated with relevant topics related to the emotion expressed.

Opinion topic identification differs from topic segmentation [3]. The opinion topics are not necessarily spatially coherent as there may be two opinions in the same sentence on different topics, as well as opinions that are on the same topic separated by opinions that do not share that topic [20]. The hypothesis is established by applying the technique of co-reference classification for topic annotation. The building of fine-grained topic knowledge based on rhetorical structure and segmentation of topics using heuristic classification substantially reduces the problem of opinion topic distinction.

The knowledge of Rhetorical Structure Theory in the text structure was used in [22] to improve the identification of topical words in a text document. The similarity between a text and its title is used to identify the text structure. The work is done at document level and not aimed for opinion topic. But, it has shown that the use of rhetorical structure knowledge can effectively identify the text structure.

The use of rhetorical structure to discover relations among topical entities that appear in the target span is the contribution for the topic identification task. Clustering the intervening context considers *Named Entities* (NE), heads of proper noun phrases and the common relations held between NPs ([8],[18]). Apart from *Named Entities* (NE), the technique used for heuristic classification is aimed at capturing the context with the help of *Emotion Holder* (EH) and four types of *Similarity* features. This technique is also applied to identify the relevant topics among all probable multiple topics.

3. Emotion Topic Annotation

One of the major problems of emotion topic extraction is the lack of appropriately annotated corpora. The blog corpus [19] tagged with any of the Ekman's six emotion types at sentence level is used in our present task. As an individual emotion topic consists of single word or a string of successive words, the annotation task is conducted to identify the scope of the topic spans in a sentence. Three annotators presented as A1, A2 and A3 have used an open source graphical tool [1] to carry out the annotation of topic spans in each of the blog sentences. Although the annotation is being carried out on the whole corpus, we have considered only 500 sentences as our test corpus in the present task. To accomplish the goal, we have used two standard metrics for measuring inter-annotator agreement.

Firstly, we have used Cohen's *kappa* coefficient (κ) [28]. It is a statistical measure of inter-rater agreement for qualitative (categorical) items. It measures the agreement between two raters who separately classify items into some mutually exclusive categories.

Secondly, we have chosen the measure of agreement on set-valued items (MASI) [29] that was used for measuring agreement on co reference annotation [29] and in the evaluation of automatic summarization [30]. MASI is a distance between sets whose value is 1 for identical sets, and 0 for disjoint sets. For sets A and B it is defined as:

MASI = J * M, where the Jaccard metric (*J*) is

$$J = |A \cap B| / |A \cup B|$$

Monotonicity (*M*) is defined as,

$$1, if A = B$$
$$2/3, if A \subset B or B \subset A$$
$$1/3, if A \cap B \neq \phi, A - B \neq \phi, and B - A \neq \phi$$
$$0, if A \cap B = \phi$$

Both of the agreement strategies assume the whole sentence as the target span. It is observed that in both strategies, the agreement for annotating target span is (≈ 0.9) that signifies almost satisfactory annotation. But, the disagreement occurs in topic level annotation. In case of emotion topic annotation, the selection of emotion topic from other relevant topics causes the disagreement. It is observed that the average number of emotion topics in sentences containing multiple topics is 2~3. The inter-annotator agreement results of the two strategies per emotion class are shown in Table 1. The low agreements in topic annotation show the problem in identifying the lexical scopes or spans for each of the emotion topics in a sentence. The agreement in identifying emotion topics in emotional sentences containing single emotion topic is more than the agreement in identifying emotion topics in sentences containing multiple emotion topics. It is decided to form the gold standard set if at least two out of three annotations matches in case of Kappa or MASI.

4. Baseline Model

The baseline model is based on the *Object* information present in the dependency relations of parsed emotional sentences. Stanford Parser [15], a probabilistic lexicalized parser containing 45 different part of speech (POS) tags of Pen Tree bank is used to get the parsed sentences with dependency relations. The dependency relationships extracted from the parsed data are checked for the predicates "*dobj*" so that the *object* related

[1] http://gate.ac.uk/gate/doc/releases.html

information in the "*dobj*" predicate is considered as the probable candidate for emotion topic. It is observed that only the *dobj* based dependency relations fail to capture the topic spans inscribed in a text. Thus one rule based argument structure acquisition framework is added to the baseline system.

Table 1. Inter-annotator Agreement using *Kappa* and MASI

Emotion Class (#Sentences) [#Topics]	Agreement measures between Pair of annotators (Kappa, κ) *[MASI]*			
	A1-A2	A2-A3	A1-A3	Avg.
Happy (92) [158]	(0.67) *[0.64]*	(0.61) *[0.60]*	(0.69) *[0.68]*	(0.65) *[0.64]*
Sad (88) [144]	(0.63) *[0.62]*	(0.67) *[0.65]*	(0.64) *[0.63]*	(0.64) *[0.63]*
Anger (84) [96]	(0.69) *[0.65]*	(0.62) *[0.61]*	(0.64) *[0.62]*	(0.65) *[0.62]*
Disgust (75) [83]	(0.67) *[0.64]*	(0.66) *[0.62]*	(0.68) *[0.64]*	(0.67) *[0.63]*
Fear (77) [84]	(0.68) *[0.67]*	(0.67) *[0.65]*	(0.69) *[0.66]*	(0.68) *[0.66]*
Surprise (84) [106]	(0.66) *[0.64]*	(0.65) *[0.63]*	(0.62) *[0.61]*	(0.64) *[0.62]*

The verb specific syntactic argument structure or subcategorization information plays an important role to identify the emotion topic. The approach is related to some earlier works ([9], [6]) done for emotion holder and topic identification using VerbNet information.

VerbNet [10] associates the semantics of a verb with its syntactic frames and combines traditional lexical semantic information such as *Thematic Roles*, semantic predicates, with syntactic frames and selectional restrictions. Verb members in the same VerbNet class share common syntactic frames, and thus they are believed to have the same syntactic behaviour. The VerbNet files containing verbs with their possible subcategorization frames and membership information are stored in XML file format. Hence, the XML files of the VerbNet are pre-processed to build up a general list that contains all member verbs and their available syntactic frames with *topic* related *thematic* information (e.g. *Topic*, *Theme*, *Event* etc.). The pre-processed list is searched to acquire the syntactical frames for each verb.

On the other hand, the parsed emotional sentences are passed through a rule based *phrasal-head* extraction module to identify the phrase level argument structure of the sentences with respect to their verbs. The extracted *head part* of every phrase from the well-structured bracketed parsed data is considered as the component of the argument structure. The acquired argument structures are compared against the extracted VerbNet frame syntaxes. If the acquired argument structure matches with any of the extracted VerbNet frame syntaxes, the emotion topic corresponding to each verb is tagged in the appropriate slot of the acquired argument structure. The topic related information such as Topic, Theme, Event etc. as specified in the VerbNet is properly tagged in the correct position of the sentences.

For Example 1, the Parse tree, dependency relations, acquired argument structure and VerbNet Frame Syntax for the verb *cry* are as follows.

<u>Parse Tree:</u> (ROOT (S (**NP** (PRP He))(ADVP (RB first))(**VP** (VBD cried)(PRT (RP up)) (**NP** (DT the)(NN toy)(NN car)))(. .)))

<u>Dependency Relations:</u> nsubj(cry-3, He-1), advmod(cry-3, first-2), prt(cry-3, up-4), det(car-7, the-5), nn(car-7, toy-6), ***dobj***(**cry**-3, **car**-7)

Acquired Argument Structure: [NP VP NP]

Simplified Extracted VerbNet Frame Syntax: [<NP value="*Agent*" ></VERB><NP-***topic***>]

The baseline considers the predicate ***dobj***(**cry**-3, **car**-7) whereas the *phrasal heads* are extracted from the parse tree to form the argument structure. As the acquired argument structure matches with the extracted VerbNet frame syntax, the phrase associated in the topic slot of the VerbNet frame syntax is mapped to the corresponding phrase in the argument structure. The phrase is therefore tagged as potential topic span in the sentence.

5. Hybrid Approach

5.1. Rhetorical Structure Extraction (RSE)

Instead of identifying rhetorical relations, the main focus of the present task is to automatically acquire the rhetorical components such as locus, nucleus and satellite from a sentence as these rhetoric clues help in identifying the individual topic spans associated in a target span of sentence. The topic of an opinion depends on the context in which its associated opinion expression occurs [21]. Hence, the primary assumption that is considered to identify emotion topic from emotional sentences is stated as follows. The part of the text span containing emotional expression (word/phrase) is considered as locus. As the locus presents in nucleus or satellite, the text span containing both nucleus and satellite except locus is considered as our primary target span that contain the potential emotion topics. The separation of nucleus from satellite is followed by the identification of locus, emotion holder and target span.

Primarily, the separation of *nucleus* from *satellite* is done based on the punctuation markers such as comma (,) (!) (?) or a *causal* keyword list containing 32 keywords (*as, because, that, while, whether etc*). But, it is observed that the *discourse markers* and *causal* verbs are also the useful clues if they are explicitly specified in text. The identification of *discourse markers* from written text itself is a research area. Hence, the present task aims to identify only the explicit *discourse markers* that are tagged by

*conjunctive_*() or *mark_*() type dependency relations of the parsed constituents. The dependency relations containing *conjunctive* markers (*conj_and*(), *conj_or*(), *conj_but*()) are considered for separating *nucleus* from *satellite* if the markers are present in between two successive clauses that are tagged as *S* or *SBAR* in the output of the parse tree. Otherwise, the component contained in *mark_*() type dependency relation is considered as a *discourse marker*. In Example 3 and Example 4, the separation of *nucleus* and *satellite* is done based on *causal* keyword "*because*".

The list of *causal* verbs is prepared by processing the XML files of English VerbNet [10]. If any VerbNet class file contains any frame with semantic type as *Cause,* we collect the member verbs of that XML class file and termed the member verbs as causal verbs. If any clause tagged as *S* or *SBAR* in the parse tree contain any *causal* verb, the clause is considered as the *nucleus* and the rest of the clauses are denoted as *satellite.* The list contains a total number of 250 *causal* verbs. (e.g. "{They **cause** tears to run down my cheeks} [that in turn make me want to fall to my knees.]").

It is to be mentioned that, the discourse markers, causal keywords or causal verbs not only separate the clauses but also the phrases of a single sentence into nucleus and satellite ("{I feel really alone right now} [because it's Friday.]"). Each word of a sentence is searched through WordNet Affect [2] and if any match occurs, the word is referenced as the locus point.

By definition, emotion topic is intended by its holder [20]. Hence, the knowledge of emotion *holder* is incorporated to extract the target span. Each of the sentences is passed through a holder identification module [6]. This syntax-based module gives *F-Score* of 66.98% on ~5K example sentences of VerbNet [10] where the example sentences contain emotion verbs that are retrieved from *WordNet Affect* [2]. The phrase or text span responsible for emotion holder is identified and tagged in the sentences.

The text span containing *nucleus* and *satellite* except the *locus* and emotion *holder*, termed as Maximum Target Span (*Max_TS*) is filtered from each sentence. The maximum possible target span contains its associated emotion topics. But, our goal is to identify more focused and fine-grained target span so that the topic spans contain less adjunct components and are identified easily. (e.g. "***Max ignored*** *the issues of sports as well as politics*").

Each of the parsed constituents of *Max_TS* is passed through a sub-module to identify any clue that helps to synchronize the target span. If any word of *Max_TS* co-exists with *locus* and emotion *holder* in the *direct* or *transitive* dependency relations, the words are allowed to construct the target span. The *direct* dependency is identified based on the simultaneous presence of both the *locus/holder* and the word in the same dependency relation whereas the *transitive* dependencies are verified if the *locus/holder* and the word are connected via one or more intermediate dependency relations. It is observed that the *transitive* dependency performs well in case of identifying target span from *satellite* part whereas the direct dependency helps in identifying target span from *nucleus* or single clause. This task helps in identifying the focused target span that is termed as *Max_FTS*. But, the final identification of each and individual emotion topic span from the entire target span is carried out using a *Heuristic Classifier* (HC).

5.2. Heuristic Classifier (HC)

Each of the topic spans either contains a single word or a string of words. Hence, the topic spans containing phrases require chunking of the input sentences.

Firstly, the emotional sentences are tagged with an open source Stanford Maximum Entropy based POS tagger [4]. The best reported accuracy for the POS tagger on the Penn Treebank is 96.86% overall and 86.91% on previously unseen words.

Secondly, the POS tagged sentences are passed through a Conditional Random Field (CRF) based chunker [26] to acquire the chunked data where each word of the chunked sentences is marked either with *beginning* (B) or *intermediate* (I) or *end* (E) tags. The word elements from the *beginning, intermediate* and *end* tagged components are appended to construct the respective phrases. Each of the chunked phrases present in the focused target span i.e., in *Max_FTS* is considered as the probable emotion topics. The chunking is also aimed for capturing the longer topic spans as the topic spans contain multiple word tokens.

Thus lastly, an unsupervised *Heuristic Classifier* (HC) is designed to assign a heuristic score (*Hscore*) to each of the chunked phrases of the *Max_FTS*. This score is assigned to classify and identify the topic related chunked phrases from the *Max_FTS*. The heuristic score (*Hscore*) is calculated by cumulating the feature scores (*FetScore*) of each identified feature. If there is *l* number of features identified for a chunked phrase, the equation of calculating *Hscore* for that phrase is as follows.

$$Hscore = \sum (Fetscore * l)$$

The *Fetscore* value is considered as a fixed decimal value with respect to all features. By reviewing the blog corpus, the following heuristic features are selected.

Emotion Holder (EH): The emotion holder identification module described in [6] is used in the present task to annotate the emotion *holders* in the blog sentences. The emotion *holder* information not only aims to identify the focused target span but also contributes in heuristic classification technique. If any *direct* or *transitive* dependency relation holds between any word element of a chunked phrase and the emotion *holder*, the

whole chunked phrase is considered as the topic span. (Example 1, the relations are *nsubj(cry-3, **He**-1), dobj(cry-3, **car**-7), nn(**car**-7, **toy**-6)* that contain a chain from emotion holder "*he*" to topic "toy car". This feature shows the usefulness in case sentences containing multiple clauses and long distance dependencies.

Named Entity (NE): Each of the sentences is passed through a Stanford Named Entity Recognizer [2] for identifying the named entities. If any word of a chunked phrase in the focused target span is tagged as a named entity, the phrase is selected as a potential topic and a feature score (*Fetscore*) is added for calculating the heuristic score (*Hscore*) for that phrase (e.g. "{*I forgot*} [*how demeaning **BME** classes are.*]").

Structural Similarity (StrucSim): If any word element of a chunked phrase and the *locus* co-occur in the *nucleus* or in the *satellite,* the feature is considered as *common similarity* whereas if they occur separately in the *nucleus* and the *satellite*, the feature is considered as *distinctive similarity*. Feature score (*Fetscore*) of zero is assigned for *distinctive similarity* (e.g. "{I ***enjoyed*** the **summer vacation**}[…]").

Sentiment Similarity (SentiSim): The *positive* or *negative* valence of each word (*pretty*, *good*) in a chunked phrase is measured using *SentiWordNet* [1]. If the words contained in the chunked phrase are present in the *SentiWordNet,* the corresponding feature entails that the phrase contains *Sentiment Similarity.* In addition to that, if the word contains either *positive* or *negative* sentiment score (> 0.0), an extra feature score (*Fetscore*) is assigned to that phrase (e.g. "*overall it was a **pretty** **good** tournament*").

Syntactic Similarity (SynSim): The syntactic similarity feature is identified with the help of a context window containing POS level argument structure present between the phrase and the *locus.* Only the chunked phrases containing verb, noun and preposition are considered. This feature is identified from the extracted argument structures of the sentences. The detail description of argument structure acquisition is mentioned in Section 4. If the phrase is already defined as a *theme* or *topic* or *event* in the baseline argument extraction module, the chunked phrase is then selected for emotion topic and the feature score is assigned in a similar way (e.g. "*He first cried up the **toy car***").

Semantic Similarity (SemSim): The semantic similarity is identified with the help of the WordNet features identified between any word of a chunked phrase and the *locus.* The features are defined as follows.

WordNet Synonymy: If any word of a chunked phrase belonging to the focused target span (*Max_FTS*) and the *locus* present in any synset of WordNet, the corresponding chunked phrase is considered as the probable candidate for emotion topic span (e.g. "*I **won** the financial **profit**.*").

WordNet Hypernymy: If any word of a chunked phrase belonging to the focused target span (*Max_FTS*) is defined as *event, topic, theme, subject, issue* or *matter* in its *hypernym* tree, the corresponding chunked phrase is considered as the probable candidate for emotion topic span (e.g. "*you at least suffered the **circumstances***").

WordNet SenseID: If the word and the *locus* both share at least a common *SenseID*, the corresponding chunked phrase is then selected as the candidate for emotion topic span (e.g. "*He can **enjoy** his **love** with freedom.*").

Individual feature score (*Fetscore*) for each of the semantic similarity features is assigned to that phrase. A morphological analyzer available with the English WordNet [7], is applied to identify the stem form of the words for identifying specially the *Sentiment* and *Semantic* similarities. But, assigning the responsible phrases as emotion topics shows that the article or conjunct present at the start or end point of the chunked phrases cause the error in matching the scope of the topic spans. For Example 1 and Example 2, the phrases "***the** toy car*", "***the** issues of sports*" started with the article "the" cause the problem of selecting the topic spans "toy car" and "*issues of sports*" respectively. Hence, the leading or ending articles or conjuncts of the chunked phrases are not considered for selecting the topic spans.

6. Evaluation

The performance of the system is measured using the standard Precision, Recall and F-score metrics. The systems are evaluated on 500 blog sentences annotated with single or multiple emotion topics. The baseline system, based on Object based dependency relations, has achieved an F-Score of 41.45%. The results per emotion class are shown in Table 2. The inclusion of argument structure knowledge into the baseline system improves the average F-score value to 45.47%. It has been observed that the baseline system identifies the emotion topic mostly from the sentences that contain single topic. The argument structure based baseline system suffers from capturing the argument structures especially from some unstructured blog sentences ("ok, shut up!", "OH OH OH he is yelling at me, "Are you ready to go, already!" "). The preliminary investigation suggests that the baseline system based on syntactic knowledge identifies topic spans but fails to identify the scope of the topic spans. The sentences containing multiple relevant emotion topics are not tagged by the system. In addition to that, the passive sentences are the sources of error in topic identification.

The unsupervised hybrid system identifies multiple emotion topics from a sentence. But, the selection of features is verified on a set of 200 development sentences and satisfactory results are produced for all the features.

[2] http://nlp.stanford.edu/software/CRF-NER.shtml

Hence, we retain all the features and employ the features to *Heuristic Classifier* (*HC*) for identifying topics from 500 test sentences. Instead of selecting a single topic based on the *Hscore*, the present technique allows us to select all identified topics with a threshold *Hscore* (>*.5*) as the potential emotion topics. If a sentence contains single emotion topic, the responsible phrase with best heuristic score (*Hscore*) is selected as the potential candidate of emotion topic for that sentence. But, the selection of the emotion topics is slightly relaxed by considering the phrases with next highest heuristic scores (*Hscore).* The results obtained by considering the highest and next highest heuristic scores are mentioned as *[H1]* and [H2] in Table 3. It is observed that further relaxation in the selection strategy does not significantly improve the performance of the system.

Table 2. *Recall*, *Precision* and *F-Score* of the Baseline System (in %)

Emotion Class (#Sentences) [#Topics]	Baseline [+ *Syntactic*]		
	Recall	Precision	F-Score
Happy (92) [158]	44.22 [*48.45*]	40.16 [*43.33*]	42.09 [*45.74*]
Sad (88) [144]	43.09 [*47.66*]	39.88 [*44.65*]	41.42 [*46.10*]
Anger (84) [96]	46.64 [*50.02*]	42.31 [*45.76*]	44.36 *[47.79]*
Disgust (75) [83]	37.34 [*41.05*]	35.67 [*39.44*]	36.48 *[40.22]*
Fear (77) [84]	39.24 [*43.18*]	36.28 [*40.58*]	37.70 *[41.83]*
Surprise (84) [106]	48.75 [*52.32*]	44.76 [*50.11*]	46.66 *[51.19]*

The error occurs mostly for unstructured sentences (e.g. "*Really starting to lose it*.") and the sentences containing typographic errors (e.g. "*she's feeling very <u>goooood</u> about herself*."). The boundaries of the chunked phrase causes problem of selecting the lexical scopes of the topic spans. But, it has been observed that the system outperforms the baseline for the sentences containing multiple topics. The average *recall, precision* and *F-score* of the hybrid system are 60.37%, 57.49% and 58.88% respectively. Furthermore, the hybrid approach is strong enough to identify the topic spans from a sentence. Thus, it is to be mentioned that the application of rhetorical structure gives a focused target span that inturn helps the Heuristic Classifier (*HC*) to identify the individual topic spans. As the *F-score* of the hybrid approach outperforms the baseline system using relaxed scoring, it signifies the fact that the performance of the hybrid approach is compatible enough to capture multiple emotion topics along with the topic spans.

Table 3. *Recall*, *Precision* and *F-Score* of the Hybrid Approach (in %)

Emotion Class (#Sentences) [#Topics]	(Identifying Topic Span) *[H1]* [H2]		
	Recall	Precision	F-Score
Happy (92) [158]	*[56.75]* 64.92	*[52.88]* 60.07	*[54.74]* 62.40
Sad (88) [144]	*[52.43]* 61.07	*[50.03]* 58.67	*[51.20]* 59.84
Anger (84) [96]	*[55.80]* 62.87	*[53.41]* 59.44	*[54.57]* 61.10
Disgust (75) [83]	*[50.84]* 54.17	*[48.04]* 52.92	*[49.40]* 53.53
Fear (77) [84]	*[49.20]* 53.97	*[44.05]* 50.48	*[46.48]* 52.16
Surprise (84) [106]	*[57.11]* 65.27	*[54.86]* 63.37	*[55.96]* 64.30

7. Conclusions

The present task identifies emotion topic from blog texts. The identification of focused target span by Rhetorical Structure and the topic span by heuristic classifier is an unsupervised but a novel contribution towards emotion topic annotation. As we have incorporated the knowledge of emotion holder, our future task is to explore the tracking of topics that are associated with the same or different emotion holder. The topic level annotation is being carried out on a large corpus to explore the current task using fuzzy or supervised technique. The inclusion of rhetorical relations for identifying topic spans is aimed for future applications.

References

[1] Esuli Andrea, and Fabrizio Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining", LREC-06, 2006.

[2] Strapparava Carlo, and Alessandro Valitutti, "WordNet-Affect: an affective extension of WordNet". In Proceedings of the 4th International Conference on Language Resources and Evaluation, pp. 1083-1086, 2004.

[3] Choi, F., "Advances in domain independent linear text segmentation". Proceedings of NAACL, 2000.

[4] Manning Christopher D., and Kristina Toutanova, "Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger", *SIGDAT Conference on Empirical Methods (EMNLP/VLC)*, 2000.

[5] Baker Collin F., Charles J. Fillmore, and John B. Lowe, "The Berkeley FrameNet project", COLING/ACL-98, pp. 86-90, Montreal, 1998.

[6] Das D., and S.Bandyopadhyay, "Emotion Holder for Emotional Verbs – The role of Subject and Syntax", CICLing- 2010, A. Gelbukh (Ed.), LNCS 6008, pp. 385-393, Romania, 2010.
[7] Miller George A., "WordNet: An on-line lexical database", International Journal of Lexicography", vol. 3(4), pp. 235–312, 1990.
[8] Hasegawa T., S. Sekine, and R. Grishman, "Discovering relations among named entities from large corpora", In Proceedings of ACL, 2004.
[9] Kim S., and E. Hovy, "Extracting opinions, opinion holders, and topics expressed in online news media text", In Proceedings of ACL/COLING Workshop on Sentiment and Subjectivity in Text, 2006.
[10] Kipper-Schuler K., "VerbNet: A broad-coverage, comprehensive verb lexicon". Ph.D. thesis, Computer and Information Science Dept., University of Pennsylvania, Philadelphia, PA, 2004.
[11] Kobayashi N., K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima, "Collecting evaluative expressions for opinion extraction", In Proceedings of IJCNLP, 2004.
[12] Azar M., "Argumentative Text as Rhetorical Structure: An Application of Rhetorical Structure Theory", Argumentation, vol. 13, pp. 97–114, 1999.
[13] Mann, W., and S. Thompson, "Rhetorical Structure Theory: Description and Construction of Text Structure", In G. Kempen (ed.), Natural Language Generation, Martinus Nijhoff, The Hague, pp. 85–96, 1987.
[14] Mann, W. C., and S. A. Thompson, "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization", TEXT 8, pp. 243–281, 1988.
[15] Marneffe Marie-Catherine de, Bill MacCartney, and Christopher D.Manning., "Generating Typed Dependency Parses from Phrase Structure Parses", 5th International Conference on Language Resources and Evaluation, 2006.
[16] Ekman Paul, "Facial expression and emotion". American Psychologist, vol. 48(4), pp. 384–392, 1993.
[17] Popescu, A., and O. Etzioni, "Extracting product features and opinions from reviews", In Proceedings of HLT/EMNLP, 2005.
[18] Rosenfeld, B. and R. Feldman, "Clustering for unsupervised relation identification", In Proceedings of CIKM, 2007.
[19] Aman Saima, and Stan Szpakowicz., "Identifying Expressions of Emotion in Text". V. Matoušek and P. Mautner (Eds.):TSD 2007, LNAI 4629, pp. 196–205, 2007.
[20] Stoyanov, V., and C. Cardie, "Annotating topics of opinions", In Proceedings of LREC, 2008.
[21] Stoyanov V., and C. Cardie, "Topic Identification for Fine-Grained Opinion Analysis", Coling 2008, pp. 817–824, 2008.
[22] Nomoto Tadashi, and Yuji Matsumoto, "Exploiting Text Structure for Topic Identification", Proceedings of the 4th Workshop on Very Large Corpora, pp.101-112, 1996.
[23] Theresa Wilson, Personal communications, 2005.
[24] Wiebe, J., T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language", Language Resources and Evaluation, vol. 1(2), 2005.
[25] Wiebe, J., Personal communication, 2005.
[26] Phan Xuan-Hieu., "CRFChunker: CRF English Phrase Chunker", PACLIC, 2006.
[27] Yi J., T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques", In Proceedings of ICDM, 2003.
[28] Cohen J., "A coefficient of agreement for nominal scales", Educational and Psychological Measurement, vol. 20, pp. 37–46, 1960.
[29] Passonneau R. J, "Computing reliability for coreference annotation", In Proc. International Conf. on Language Resources and Evaluation, Lisbon, 2004.
[30] Passonneau R. J., "Measuring agreement on set-valued items (MASI) for semantic and pragmatic annotation", In Proc. 5th Int'l Conf. on Language Resources and Evaluation, 2006.