

Recognition of Partial Textual Entailment for Indian Social Media Text

Dwijen Rudrapal¹, Amitava Das², Baby Bhattacharya¹

¹ National Institute of Technology, Agartala,
India

² Indian Institute of Information Technology, Sricity, Andhra Pradesh,
India

dwijen.rudrapal@gmail.com, amitava.santu@gmail.com, babybhatt75@gmail.com

Abstract. Textual entailment (TE) is a unidirectional relationship between two expressions where the meaning of one expression called *Hypothesis (H)*, infers from the other expression called *Text (T)*. The definition of TE is rigid in a sense that if the H entails from T but lacks minor information or have some additional information, then the pair is treated as non-entailed. In such cases, we could not measure the relatedness of a *T-H* pair. Partial textual entailment (PTE) is a possible solution of this problem which defines partial entailment relation between a *T-H* pair. PTE relationship can play an important role in different Natural Language Processing (NLP) applications like text summarization and question-answering system by reducing redundant information. In this paper we investigate the idea of PTE for Indian social media text (SMT). We developed a PTE annotated corpus for Bengali tweets and proposed a Sequential Minimal Optimization (SMO) based PTE recognition approach. We also evaluated our proposed approach through experiment results.

Keywords. Textual entailment, social media text, text summarization, partial textual entailment, question-answering system, machine learning.

1 Introduction

Recognition of textual entailment (RTE) [6] is an important task in the study of Natural Language Processing (NLP). Important role of RTE includes question-answering system where entailment status verified for provided answer in respect of expected answer, multi-document summarization where entailed sentences should be removed, and information retrieval system where only query

entailed documents are desirable. However, the classical definition of TE holds two important constraints over the relation. One, the relation is unidirectional where *H* must entail from *T*. Else the pair is non-entailed.

Still, in many NLP applications like question answering system or text summarization, both the conditions are equally important to reduce redundant information from the final text. Two, the definition of TE is firm in respect of expressions boundaries such that, the complete meaning of *H* must be inferred from complete meaning of *T*. It implies that, if *H* entails from *T* and has some additional information or lacks of some information then the relation is a false TE. Since RTE is a binary decision problem, in case of false TE relation, it is quite difficult to decide whether the relation is completely negative or mostly positive. For example, in the following *T-H* pair, *H* entails from *T* but lacks of some information. So, as per the TE definition, the cited pair is non-entailed.

T : স্কটল্যান্ডের স্বাধীনতা ভোটে হেরে (পদত্যাগ করছেন স্যামন্ড)

Eng. Gloss: *Samond resigned after defeated in Scotland's Independence election.*

H : ফাস্ট মিনিস্টার স্যালমন্ডের পদত্যাগ

Eng. Gloss: *First minister Samond resigned.*

The possible solution of these constraints is partial textual entailment (PTE) relationship which extends TE with partial relation. The work [14] first introduced the concept of partial meaning inferring into texts. Later the idea of PTE introduced by

the work [11] where PTE is explicitly defined and explained under the faceted entailment model.

In this paper, our contributions are as follows:

1. We extend PTE by introducing two categories to define partially entailed relation besides classical TE categories. We investigate PTE relation for Indian social media text specifically for Bengali tweets and propose a PTE recognition (RPTE) system based on machine learning algorithm. Our motivation behind the SMT is that the recent interest of the research community as well as industry on SMT such as Twitter in the last few years. Processing of social media text is more challenging than carefully authored traditional text like news text due to their short, noisy, context dependent and dynamic nature. We justified the performance of the proposed approach through experiment results.
2. We developed a corpus of Bengali tweet pairs and manually annotated by PTE classes. The corpus includes total 5916 numbers of tweet pairs.

The rest of the paper is organized as follows. In the next sections, we discuss the related research work in TE and PTE domain followed by section 3 which formalizes partial entailment categories with appropriate examples. Section 4 reports corpus preparation and section 5 reports proposed approach. Section 6 explains experiment and result followed by the a result analysis in section 7. Finally, we draw the conclusion of our research with future scope of the work in Section 8.

2 Related Work

Our research work is closely related to textual entailment. So, in this section we discuss related and promising research works on PTE as well as TE.

The work [14] first introduced the concept of partial meaning inferring into texts. In the proposed work the reference answer is break down into an ordered pair of words (w_1 , w_2) that are contained in the hypothesis accompanied by a semantic relation. Ordered pair of words

is called facet which refer to some part of a text meaning. Then based on the facet overlap between student answers and reference answer every pair is tagged by one of 8 partial meaning labels instead of classical TE decision in the form of yes or no. The annotation labels are: Assumed (facet that are assumed), Expressed (Facet that are directly expressed), Inferred (Facet inferred), Contra-Expr (Facet directly contradicted by negation), Contra-Infr (Facet contradicted by pragmatics), Self-Contra (Facet that are contradicted and implied), Diff-Arg (Facet where core relation expressed) and Un-addressed (Facet not addressed at all).

Partial meaning inferring into text is extends into PTE notion in the work proposed in [11]. Hypothesis breaks down into components and recognizes whether any individual component is entailed by text or non-entailed. The decision of partial entailment is represented by three labels: Exact Match, Lexical Inference, and Syntactic Inference. System labels exact match based on the overlapping of lemmas in the pair. Lexical Inference is based on semantic relatedness of the words in facet and text measured by using WordNet based Resnik Similarity [15]. Whereas, the Syntactic Inference label depends on dependence tree of a given facet, obtained by using Lowest Common Ancestor (LCA) of facet-nodes. The work [9] and [16] have explored PTE for traditional news text and social media text. The work formalizes 4 classes of partial textual entailment and proposes a PTE recognition system based on cosine similarity measure. These research work concludes through experiments that the PTE exploration is a complex task and becomes more complex for SMT than traditional text due to the nature of text.

Vita et al. proposed a PTE recognition system [24] based on Word2vec representation under the faceted entailment model. The system defines two class of partial textual relationship like expressed and un-addressed by mentioning semantic relationship between facet and text.

RTE task is a well defined problem in NLP domain. Since the inception of automatic RTE task [6], different RTE challenges (RTE-1 to RTE-7) have been conducted on different RTE

dataset and proposed many state-of-the-art RTE systems. The systems are ranging from lexical similarity methods to complex and extensive linguistic analysis methods including creation of an entailment dataset for heterogeneous domain as well as highly competitive entailment systems.

Lexical-syntactic similarity based approaches use features like n-gram, word similarity, synonym and antonyms [10], word overlap [1] of T and H to determine the existence of entailment. In the work [25] relatedness between T and H measured for RTE task using dependency graphs and Lexical chains. The system proposed by [22], form sequence of transformations for T into H to make resulting T identical to H and decides entailment relation. Various machine learning approaches also introduced to solve RTE problem. Those approaches use various features like WordNet and numeric expressions, syntactic and semantic features [2] and also use different dataset like clinical text corpus [18]. A feature-rich classifier model and a neural network model for TE recognition is proposed by Bowman et al. [4] and evaluated both the model on the Stanford Natural Language Inference (SNLI) corpus. A very Recent work [23] proposed neural network based approach using word embeddings to recognize TE for tweets. The work evaluated the approach by comparing the performance with previous promising machine learning based approaches. Authors report the results for a batch size of 50 over 100 epochs and reports comparable results. Another approach [21] recognizes TE using distributional semantics based navigation algorithm. The approach built a knowledge graph composed of dictionary definitions of terms extracted from a linguistic resource following the work proposed by Silva et al. [20]. The navigation algorithm finds paths between the text and hypothesis in knowledge graph to take decision of entailment.

Our current research work is related to the work proposed in [9, 16] respectively. There are two substantial differences between these works and ours. Firstly, earlier works focused only on cosine similarity score between T and H . It is worth noting that, cosine similarity measure does not cover semantic meaning of texts. So, we

added more features like semantic similarity and word to vector similarity in our approach for a robust PTE recognition system. Secondly, we focus on machine learning based approach rather statistical approach followed in the earlier proposed approaches.

3 Empirical Definition of Partial Textual Entailment

In this section we define 4 categories of proposed PTE (PTE-I, PTE-II, PTE-III and PTE-IV). Each category is explained along with appropriate examples. During categorization, we break down each T and H into two sections like X_T, Y_T and X_H, Y_H respectively to explain partial entailed portion and additional information. We also use braces to identify the boundary of the entailed portion in T and H .

- **PTE-I:** If the complete meaning of H inferred from complete meaning of T or infers to complete meaning of T then it is a category of PTE-I. This category is the perseverance of the original entailment definition and the relationship is bi-directional. The relation can be expressed as:

H Entails from T

T Entails from H

For example,

T : পাস হলো সংবিধানের ১৬ তম সংশোধন বিল বিচারপতিদের অপসারণের ক্ষমতা নিল সংসদ

Eng. Gloss: 16Th correction bill of constitution has passed, Parliament holds judge removal power.

H : বিচারক অপসারণে সংবিধান সংশোধনী বিল পাস

Eng. Gloss: Constitution correction bill passed to remove judge.

- **PTE-II:** This category of PTE contains two conditions, like:

Condition I: If H entails from the whole meaning of T and have additional information, then it is a category of PTE-II and represents as,

$(X_H \text{ Entails from } T) + Y_H$

For example,

T : কাশ্মীরে ৬০ বছরের মধ্যে সবচেয়ে ভয়াবহ বন্যা

Eng. Gloss: *Kashmir faced the most devastating flood in last 60 years.*

H : (ভারতের জম্মু কাশ্মীরের ভয়াবহ বন্যা পরিস্থিতি) মোকাবিলা করতে আজ থেকে হেলিকপ্টার ও বিমানের মাধ্যমে উদ্ধারকার্য শুরু

Eng. Gloss: *Rescue operation by Helicopter and flights started today to face (the most devastating flood condition in Kashmir, India).*

Condition II: If T entails from the whole meaning of H and have additional information, then it is also a category of PTE-II and represents as,

$(X_T \text{ Entails from } H) + Y_T$

For example,

T : (যুদ্ধবিরতির মধ্যেই পাল্টাপাল্ট গোলাবর্ষণ) পূর্ব ইউক্রেনের বন্দর নগরী মারিওপ

Eng. Gloss: *(Cross fire under ceasefire) in Mariup city of east Ukraine.*

H : যুদ্ধবিরতির মধ্যেই ইউক্রেনে পাল্টাপাল্ট গোলাবর্ষণ

Eng. Gloss: *Cross fire under ceasefire in Ukraine.*

- **PTE-III:** If a portion of H entails from a portion of T or vice versa, then it is a category of PTE-III and represents as,

$(X_H \text{ Entails from } T) + Y_H$

$(X_T \text{ Entails from } H) + Y_T$

For example,

T : (রাশিয়াপন্থী বিচ্ছিন্নতাবাদী ও সরকারি বাহিনীর পাল্টাপাল্ট গোলাবর্ষণ ও সংঘর্ষের ঘটনায় ইউক্রেনে চলমান যুদ্ধবিরতি) নিহত ১ নাগরিক

Eng. Gloss: *(Russia supported separatist and government army are in open crossfire under Ukraine ceasefire), one citizen death reported.*

H : (ইউক্রেনে পাল্টাপাল্ট গোলাবর্ষণ অস্ত্রবিরতি সত্ত্বেও) হুমকির মুখে ইউক্রেনের পূর্বাঞ্চলের কৌশলগত শিল্পনগর

Eng. Gloss: *(Open crossfire is going on even though ceasefire in Ukraine), industrial city at eastern Ukraine is under threat.*

- **PTE-IV:** If T or H does not entail from H or T , then its a category of PTE-IV and represents as Non-entailed.

For example,

T : সংবিধানের ষোড়শ সংশোধনীর বিরোধীতাকারীদের বিষয়ে সুপ্রিম কোর্টসহ সারাদেশের আইনজীবীদের সতর্ক থাকার আহ্বান জানিয়েছেন

Eng. Gloss: *Honorable Supreme court requested all the lawyers of the country to remain alert regarding the opposition of constitution's 16Th correction.*

H : সংবিধানের ষোড়শ সংশোধনীতে শিক্ষাবিদ বা বিশিষ্টজনের কোনো মতামত নেওয়া হবে না বলে জানিয়েছেন আইনমন্ত্রী অ্যাডভোকেট

Eng. Gloss: *The Law minister advocate told that no views from the educationists and prominent figures regarding constitution's 16th correction will be accepted.*

4 Corpus Acquisition

We have prepared a corpus of tweets for Indian social media text specifically for Bengali tweets. We collected Bengali tweets on 10 topics based on recent popular twitter keywords. Our chosen topics covered various information domains like international and national politics, sports, natural disasters, political campaigns, elections etc. In particular, for a period of one month from 1st June to 30th June, 2017, we have collected 31,000 tweets. However, the number of tweets are varies in a range from 1500/topic to 9000/topic.

We filtered out Re-tweets and very short incomplete tweets from the corpus to avoid information redundancy and meaning less tweets respectively. Every pre-processed tweet paired with all the other tweets within the topic. To make number of tweet pair precise and meaning full for annotation task, we followed an automatic filtration process before annotation task. In the following subsection, we discuss the filtration and annotation process in details up to some extent.

Table 1. Corpus statistics and IAA distribution

Category	Tweet pairs (Count)	IAA
PTE-I	906	0.86
PTE-II	644	0.59
PTE-III	550	0.45
PTE-IV	3816	0.97

4.1 Annotation and Corpus Statistics

The filtration process concentrated on selecting those pairs where possibility of getting PTE-I, PTE-II and PTE-III is higher. So, we calculated cosine similarity of each pair and discarded those having similarity score of 0.30 (in the scale of 0-1) or less. The reason behind choosing this value is that, our manual observation for all the pairs scores reveals that the pairs having score 0.30 or less are mostly non-entailed pairs. Thus, we have chosen that threshold score. After this initial filtration we prepared a corpus of 7000 tweet pairs to proceed for manual annotation task. We employed two human annotators for annotating each pair tweets by one of 4 PTE classes. Annotators are native Bengali speakers and annotated all the tweet pairs. To assess annotation agreement among the annotator, we measured Cohen Kappa coefficient [5] value for all the annotated pairs. Detailed categorical distribution of corpus with annotation agreement is reported in table 1. After complete annotation task, our corpus includes 5916 numbers of annotated tweet pairs for experiment work.

5 System Description

We have formulated RPTE problem as a multi-class classification problem where each instance consists of a pair. We adopt machine learning based approach to build the proposed RPTE system. The system identifies the class of textual entailment based on the various similarity measures of T-H pair at the lexical and shallow syntactic level.

We involved a pre-processing phase before computation of various similarity score. The first step of pre-processing is to divide the input text

into units called tokens. Each of these is a word, a number, a punctuation mark and many others. The Bengali POS-Tagger developed by [7] has been used for tokenizing the tweets. Punctuation marks can vary in different computational process. So, we strip out all the punctuation marks from the whole corpus.

5.1 Features

The features used in our classifiers are the different kind of similarity scores for each T-H pair. These features capture the relatedness between T and H. Here, we give brief description of each feature.

5.1.1 Cosine Similarity:

Cosine similarity is widely used in information retrieval to calculate the similarity between documents or sentences. Cosine similarity score represents the relatedness between two n-dimensional vectors. We considered binary vectors with values either 0 or 1 in our computation. Given two vectors of attributes A and B, the cosine similarity is calculated using the equation 1:

$$\cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \cdot \|\vec{B}\|} \quad (1)$$

Before cosine similarity calculation of each pair, we removed stop words and transformed each word into its root word. The stop/junk words as per proposed list ¹ is followed in our work. We develop a Bengali stemmer to transform each Bengali word into its root form.

5.1.2 Semantic Similarity:

Semantic similarity score between T and H is one of the most frequent feature for RTE task [13]. We modeled the semantic similarity of two texts (T,H) as a function of the semantic similarity of the constituent words in both texts. To achieve this objective, our system computed word to word similarity by using the approach proposed in the work [17]. Word similarity is computed based on Bengali WordNet proposed in [8]. Bengali

¹<http://www.isical.ac.in/utpal/resources.php>

WordNet is a lexical semantic network which holds semantic relations like synonyms and word-senses as the nodes and relations of the synonyms and word-senses are the edges of the network. Word level similarity score accumulated to find sentence level similarity using equation 2:

$$Semantic_{sim}(A, B) = 2 * \frac{\sum sim_{w_i}(A, B)}{\sum length(A, B)}, \quad (2)$$

where the summation of maximum similarity for each word is represents by $sim_{w_i}(A, B)$ and $Semantic_{sim}(A, B)$ is a function that returns semantic similarity score between two tweets. Similarity could not be computed for all words in the text and hypothesis, either because some words, such as proper nouns which do not appear in WordNet, or because some of these metrics can only be calculated if an information content value has already been calculated for the word sense.

An important point is that the similarity value is based on each of the individual word similarity values, so that the overall similarity always reflects the influence of each word and its senses. According to the semantic similarity score formulation, similarity values ranges from 0 to 1.

5.1.3 Word-to-Vector Similarity:

Vector representations of words lend themselves well to express similarity between words. We have used Word2Vec toolkit ² to determine word-to-vector similarity for each $T-H$ tweet pair. Word2Vec is an efficient implementation of deep learning techniques based on two architectures, continuous bag-of-words (CBOW) and skip-gram (SG) [12]. To measure word2vec similarity of a word, we used word embedding model for Bengali text as proposed in [3]. The trained model is having vocabulary size of 4,36,126 with 300-dimensional vectors and window size of 5.

Every word vector in T is compared with all the word vectors of H to find maximum similar word. Similarity score represents by a score in a range of 0 to 1, stating minimum and maximum similarity.

²<http://deeplearning4j.org/word2vec.html>

Word level similarity score accumulated to find tweet level similarity measure using equation 3:

$$w2v_{sim}(A, B) = \frac{\sum_{i=0}^{i=n} max_{sim}(w_i)}{N}, \quad (3)$$

where $max_{sim}(w_i)$ is the maximum similarity score of a word in T and N is the total number of unique words in $T-H$ pair.

5.1.4 Uni-gram Similarity:

Each uni-gram in T is searched for a match in the corresponding H part. The measure U_{sim} is calculated as the fraction of the T uni-grams that match in the corresponding H :

$$U_{sim} = \frac{U_m}{n} * 100, \quad (4)$$

where U_m is the maximum matching score and n is the number of unique words in both T and H .

5.1.5 Bi-gram Similarity:

Each bi-gram in T is searched for a match in the corresponding H part. The measure B_{sim} is calculated as the fraction of the text bi-grams that match in the corresponding H :

$$B_{sim} = \frac{B_m}{n} * 100, \quad (5)$$

where B_m is the maximum matching score and n is the number of unique words in both the tweets.

5.1.6 Longest Common Subsequence (LCS):

LCS of a $T-H$ pair is the longest sequence of words which is common to both the T and H :

$$LCS_{sim} = \frac{lcs_m}{n} * 100, \quad (6)$$

where lcs_m is the length of maximum LCS and n is the number of unique words in $T-H$ pair.

5.1.7 Skip-grams Similarity:

A skip-gram is any combination of n words as per occurrence order in the sentence allowing arbitrary gaps. In the present work, only 1-skip-bigrams are considered where 1-skip-bigrams are bigrams with one word gap between two words in order in a sentence. The measure 1-skip-bigram similarity SG_{sim} is defined as in equation 7:

$$SG_{sim} = \frac{sg_m}{n} * 100, \quad (7)$$

where sg_m is the maximum matching score and n is the number of unique words in both the tweets.

5.2 Classifier

Our proposed approach is based on SMO classifier to recognize PTE for Bengali tweets. We have also experimented with different promising machine learning algorithms those have shown comparative performance [2, 18] earlier in order to evaluate our proposed approach.

6 Experiment and Result

In this section we describe the experiment details along with a comparison to the existing RPTE system.

We have experimented with the four machine learning algorithms like Random forest (RF), Decision tree (DT), Logistic regression (LR) and SMO algorithms. These algorithms were selected because they are known for achieving higher performances for multi-class classification problems as per the literature study of previous research work. We have used WEKA³ machine learning platform [26] to perform our experiments.

We have evaluated the system using cross validation technique with ten folds, testing over the classifiers for four-way classification task. The results are shown in the table 2 below. Feature wise performance of the Classifiers shown in the table. First we asses accuracy performance based on cosine similarity (CS) feature, then based on CS and Lexical similarity features. Lexical similarity feature includes uni-gram, bi-gram, LCS

³<http://www.cs.waikato.ac.nz/ml/weka/>

and Skip-grams Similarity, next based on CS, lexical and semantic similarity (SS), finally based on all the features including word 2 vector similarity (W2V). Every classifier shown higher performance by turning-on all features. The best performance is achieved with SMO classifier which is reported as F-measure for PTE-I, PTE-II, PTE-III and PTE-IV as 78.1, 64.6, 75.5 and 98.3 respectively. The difference between the best and the worst classifier performance is 11.5, 9.2, 7.7 and 3.7 for PTE-I, PTE-II, PTE-III and PTE-IV respectively using all features.

We also compared our proposed approach with the approach proposed in [16] (baseline system) on our dataset. The performance result in terms of precision (pre.), recall (rec.) and F-measure is shown in table 3.

Table 2. Detail performance of the classifiers

PTE type	Features	RF	DT	LR	SMO
I	CS	72.2	63.8	69.1	76.4
	+ Lexical	72.0	64.5	70.9	77.1
	+ SS	72.8	65.5	70.4	77.6
	+ W2V	73.9	66.6	72.5	78.1
II	CS	60.0	46.4	57.5	61.4
	+ Lexical	60.1	50.7	56.9	62.9
	+ SS	60.9	53.7	57.3	63.5
	+ W2V	63.4	55.4	58.6	64.6
III	CS	72.9	65.4	66.6	74.8
	+ Lexical	72.5	66.1	68.3	75.4
	+ SS	74.6	68.8	68.2	75.7
	+ W2V	75.3	67.8	69.5	75.5
IV	CS	93.3	94.6	94.6	97.6
	+ Lexical	94.0	94.7	93.8	97.6
	+ SS	95.0	94.6	94.9	98.3
	+ W2V	95.2	94.6	94.9	98.3

7 Result Analysis

Despite the fact that the results are showing the average accuracy, we believe that the obtained results are very promising. This is because of the fact that the PTE task is more challenging than classical TE relation. More specifically, identification of PTE-II and PTE-III is more difficult

Table 3. comparative result with existing RPTE system

Approach	PTE Cat.	Pre.	Rec.	F-measure
Baseline	I	52.7	49.2	50.9
	II	36.6	39.7	38.1
	III	26.2	58.9	36.3
	IV	82.4	67.7	74.3
Proposed	I	77.2	79.0	78.1
	II	68.0	59.9	64.6
	III	78.4	72.7	75.5
	IV	96.0	99.0	98.3

than PTE-I and PTE-IV. System recognizes PTE-II and PTE-III with accuracy of F measure 64.6 and 75.5 respectively.

The reason behind this average performance may be due to the fact that the more additional information in T or H proportionally decreased similarity scores which tends to false class identification. For example:

T : যুক্তরাজ্যের সাথেই থাকছে স্কটল্যান্ড

Eng. Gloss: *Scotland will be with United Kingdom.*

H : (ব্রিটিশ যুক্তরাজ্য আপাতত ঐক্যবদ্ধ থাকছে) সব জল্পনা কল্পনার অবসানের পর অবশেষে জানা গেল স্বাধীনতার পক্ষে স্কটল্যান্ডের ভোটদাতাদের ৫৫ শতাংশ না ভোট

Eng. Gloss: *(United Kingdom is united), at the end of all prediction, finally reported that 55% voters voted against independence of Scotland.*

In above example, in H additional information is much more than entailed information. This large amount of additional information returns very poor similarity values which recognize the pair as non-entailed even though the pair is an example of PTE-II. Similar kind of identification difficulties arises for PTE-III also.

The accuracy of PTE-I and PTE-IV recognition is under the average accuracy of state-of-the-art RTE system. The main reason of this average accuracy may be due to the nature of social media text which is completely different than traditional text and pose various challenges towards processing. In many instances it is found that traditional cosine similarity score is not representing actual

closeness between expressions. Soft cosine measure [19] may represents relatedness more accurately, which is in our future scope of the current work. Our approach outperforms to identify PTE-IV with maximum accuracy of F measure 98.3 higher than the other category of PTE. This is usual as most of the non-entailed pairs similarity scores remain in a specific lower range.

Table 2 shows that the performance of the classifiers are gained when more features are added. Significant difference in the performance is found when both the syntactical and semantic features are used instead of using only syntactical features. We also found that the classifier gains more accuracy by using all the features.

8 Conclusion and Future Work

Exploring various classes of partial entailment for Indian social media text is the core contribution of this task. We proposed an automatic RPTE approach for Bengali Social media text where information is more challenging in nature due to the unique style of writing. We have conducted experiments and showed that our proposed approach outperform over the performances of previously introduced approaches.

Our future scope of work is to make the system more robust and applicable for code-mixed social media text. Unicode Bengali tweets are less noisy in nature with compare to the code-mixed tweets. So, RPTE for this text genre would be more difficult.

References

1. Adams, R., Nicolae, G., Nicolae, C., & Harabagiu, S. (2007). Textual entailment through extended lexical overlap and lexico-semantic matching. *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, ACL, pp. 119–124.
2. Agichtein, E., Askew, W., & Liu, Y. (2008). Combining lexical, syntactic, and semantic evidence for textual entailment classification. *TAC*, NIST.

3. **Alam, F., Chowdhury, S. A., & Noori, S. R. H. (2016).** Bidirectional lstms - crfs networks for bangla pos tagging. *Computer and Information Technology (ICCIT), 2016 19th International Conference on*, IEEE, pp. 377–382.
4. **Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015).** Learning natural language inference from a large annotated corpus. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, ACL, pp. 632–642.
5. **Cohen, J. (1960).** A coefficient of agreement for nominal scales. *Educational and psychological measurement*, Vol. 20, No. 1, pp. 37–46.
6. **Dagan, I. & Glickman, O. (2004).** Probabilistic Textual Entailment: Generic Applied Modeling of Language Variability. *Learning Methods for Text Understanding and Mining*.
7. **Dandapat, S., Sarkar, S., & Basu, A. (2007).** Automatic part-of-speech tagging for bengali: An approach for morphologically rich languages in a poor resource scenario. *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, Association for Computational Linguistics, pp. 221–224.
8. **Das, A. & Bandyopadhyay, S. (2010).** Semanticnet-perception of human pragmatics. *Proceedings of the 2nd Workshop on Cognitive Aspects of the Lexicon*, pp. 2–11.
9. **Das, A. & Rudrapal, D. (2014).** Exploring the partial textual entailment problem for bengali news texts. *Research in Computing Science*, volume 86, RCS, pp. 43–52.
10. **Herrera, J., Peñas, A., & Verdejo, F. (2006).** Textual Entailment Recognition Based on Dependency Analysis and WordNet. *MLCW 2005*, pp. 231–239.
11. **Levy, O., Zesch, T., Dagan, I., & Gurevych, I. (2013).** Recognizing partial textual entailment. *ACL (2)*, pp. 451–455.
12. **Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013).** Efficient estimation of word representations in vector space. *CoRR*, Vol. abs/1301.3781.
13. **Neel, A., Garzon, M. H., & Rus, V. (2008).** A semantic method for textual entailment. *Proceedings of the Twenty-First International Florida Artificial Intelligence Research Society Conference, USA*, pp. 171–176.
14. **Nielsen, R. d., Ward, W., & Martin, J. h. (2009).** Recognizing entailment in intelligent tutoring systems*. *Nat. Lang. Eng.*, Vol. 15, No. 4, pp. 479–501.
15. **Resnik, P. (1995).** Using information content to evaluate semantic similarity in a taxonomy. *CoRR*, Vol. abs/cmp-lg/9511007.
16. **Rudrapal, D. & Bhattacharya, B. (2014).** Recognition of partial textual entailment for bengali tweets. *In the Proceedings of 1st Workshop on Language Technologies for Indian Social Media (SOCIAL-INDIA), Eleventh International Conference on Natural Language Processing (ICON)*, volume 11, ICON, pp. 29–35.
17. **Rudrapal, D., Das, A., & Bhattacharya, B. (2015).** Measuring semantic similarity for bengali tweets using wordnet. *Recent Advances in Natural Language Processing (RANLP-2015)*, volume 10, RANLP, pp. 537–544.
18. **Shivade, C. P., Raghavan, P., & Patwardhan, S. (2016).** Addressing limited data for textual entailment across domains. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, August 7-12, Germany*, volume 1, pp. 1245–1255.
19. **Sidorov, G., Gelbukh, A., Gómez-Adorno, H., & Pinto, D. (2014).** Soft similarity and soft cosine measure: Similarity of features in vector space model. *Computación y Sistemas*, Vol. 18, No. 3, pp. 491–504.
20. **Silva, V. D. S., Handschuh, S., & Freitas, A. (2016).** Categorization of semantic roles for dictionary definitions. *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon, CogALex@COLING 2016, Osaka, Japan, December 12, 2016*, pp. 176–184.
21. **Silva, V. D. S., Handschuh, S., & Freitas, A. (2018).** Recognizing and justifying text entailment through distributional navigation on definition graphs. *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 4913–4920.
22. **Stern, A. & Dagan, I. (2014).** The biutee research platform for transformation-based textual entailment recognition. *Journal Linguistics Issues in Language Technology (LiLT)*, volume 9, pp. 1–26.
23. **Sulea, O. (2017).** Recognizing textual entailment in twitter using word embeddings. *Proceedings of the 2nd Workshop on Evaluating Vector Space Representations for NLP, RepEval@EMNLP 2017*,

Copenhagen, Denmark, September 8, 2017, pp. 31–35.

24. **Vita, M. & Križ, V. (2016)**. Word2vec based system for recognizing partial textual entailment. *Computer Science and Information Systems (FedCSIS), 2016 Federated Conference on*, IEEE, pp. 513–516.
25. **Wang, R. & Neumann, G. (2007)**. Recognizing Textual Entailment Using Sentence Similarity based on Dependency Tree Skeletons. *Proceedings of the*

ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, ACL, Prague, pp. 36–41.

26. **Witten, I. H. & Frank, E. (1999)**. Data mining: practical machine learning tools and techniques with java implementations.

Article received on 19/10/2017; accepted on 07/05/2018.
Corresponding author is Dwijen Rudrapal.