

# A New Approach for Twitter Event Summarization Based on Sentence Identification and Partial Textual Entailment

Dwijen Rudrapal<sup>1</sup>, Amitava Das<sup>2</sup>, Baby Bhattacharya<sup>1</sup>

<sup>1</sup> National Institute of Technology, Agartala, India

<sup>2</sup> Wipro AI Lab, Bangalore, India

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

**Abstract.** Recent trend of information propagation on any real-time event in Twitter makes this platform more and more popular than any other online communication media. This trend creates a necessity to understand real-time events quickly and precisely by summarizing all the relevant tweets. In this paper, we propose a two-phase summarization approach to produce abstract summary of any Twitter event. The approach first extracts key sentences from the whole set of event relevant tweets and eliminates maximum redundant information by exploring Partial Textual Entailment (PTE) relation between sentences. Next, generates an abstract summary over the least redundant key sentences. We conduct experiments to evaluate the performance of our propose approach and report that the approach outperforms over the baseline approach as well as state-of-the-art event summarization approach.

**Keywords.** Social media text, twitter event, summarization, partial textual entailment, tweet ranking.

## 1 Introduction

Summarization is a process of presenting an event's most important and relevant information in precise way. It helps to understand an event very quickly and precisely. The significance of summary becomes more imperative and effective while dealing with events in social media platforms like Twitter [34] where information is fast, rich content, diverse and ever increasing. Irrespective of any geographic location, time and other conditions, any user can post comments or views, upload videos on any event very quickly. Thus, Twitter has

become one of the most popular online information sharing platforms and source of breaking news as well. Twitter is the first to report information on many events like earthquake in North East India, terrorist attacks in German and France, missing of m327 flight, status on USA President election, status and result of sports events and many more before any traditional news media. However, the information shared in Twitter for an event often leads to the information overload problem [21, 45] leaving very less number of event relevant posts [53]. Though, the summarization task of Twitter events received a lot of attention from research world throughout the last decade due to the sharing of fast and diverse response from millions of users.

In general, summarization is a user specific task where generated summaries are varied due to the diverse views of the users concerning the event. While, summarization of Twitter event is more difficult than traditional documents due to the different nature of text genre which poses several challenges. Such as:

1. **Processing of tweet content:** Tweets are limited to 280 characters and often published without proofreading. Often, tweets content include spelling mistakes, grammatical errors, self created acronyms, out-of-vocabulary words and very short comments from online discussion of users. In addition, a large number of tweets in an event are irrelevant to that event or relevant to other event. Thus, standard Natural Language Processing (NLP) tools do not work well on this genre of text and

and creates challenges for processing of tweet contents, clustering and topic modeling task due to the high perceptive nature of features.

2. **Sentence boundary detection (SBD):** A tweet is not always a single sentence, rather, it may include multiple sentences. Sentences are essential prerequisite for any summarization task. But, SBD in tweets is very challenging task [36] due to the use of punctuation in a nonstandard manner. Sometimes, tweet includes no punctuation at all even though it includes multiple sentences. For example, first tweet of following examples includes two sentences with nonstandard punctuation as sentence end marker. Second tweet includes 3 sentences while 2nd and 3rd sentences have no any sentence end marker.

**Tweet 01:** *#USA + #Europe + #Asia must help #Africa ... #Liberia cannot cope alone with #Ebolaoutbreak <http://t.co/IXdRW4Q3BB>*

**Tweet 02:** *Modiji bi-election results show that people of Bihar and UP can not break the caste-religious bonds . They r poor they do't prefer growth*

3. **Information redundancy problem:** Events in Twitter attract large volume of tweets which makes difficult to identify the most important tweets for human too. Every event contains redundant information for an event in the form of Re-tweet or tweet having fully or partially entailed contents of another tweet at high volume. In order to reduce redundant information, identification of partially entailed information among tweets is a more difficult task than identifying re-tweet or entailed tweets. For example, two sentences are cited below from two different tweets. Second sentence contains entailed information from first sentence and additional content (highlighted). In order to reduce redundant information, these two sentences can be merged into one meaningful sentence as shown below.

**Sentence 01:** *Its #Jallikattu day at the world famous #Alanganallur.*

**Sentence 02:** *World Famous #Alanganallur #Jallikattu back after 3 years.*

**Merged Sentence:** *Its #Jallikattu day at the world famous #Alanganallur back after 3 years.*

This paper proposes a summarization approach to generate an abstractive summary of an event from Twitter by providing potential solution to above challenges. Figure 1 presents the structure of our summarization approach. Abstractive summarization [26, 23] characterizes an event in more compact way with the cost of high time-complexity. Our approach follows extractive-abstractive structure to generate summary over most informative and relevant sentences from entire tweet stream of an event. Experimental results in section 4 confirm that the proposed summarization approach performs better than state-of-the-art summarization approach. Our main contributions in this paper are:

1. In most of the earlier research work, a tweet is considered as a sentence for summarization task. But in reality, often tweets include multiple sentences. In this research work we extract key sentences by SBD in tweets rather than treating a tweet as a sentence. In order to extract most important sentences for an event, we ranked all the event relevant tweets first and then transformed top ranked tweets into possible sentences.
2. In order to generate abstract summary, we filter out redundant information by exploring Partial Textual Entailment (PTE) relationship among sentences. We recognize partially entailed (PE) sentences and the boundary of partially entailed text among sentences. A pair of PE sentences merged into single sentence by combining PE text. Finally, an abstract summary generates over selective sentences by covering maximum information, diversity, coherence and readability.

This paper is structured as follows. Section 2 presents current research progress of Twitter event summarization. Section 3 presents the proposed summarization approach followed by experimental

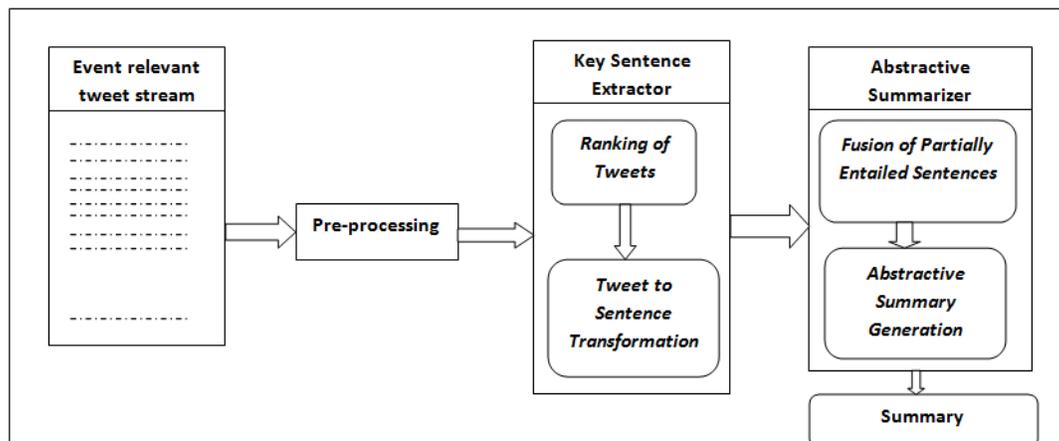


Fig. 1. Structure of Proposed summarization approach

setup and results in section 4. Section 5 reports analysis of result and finally, section 6 concludes this paper.

## 2 Related Work

Our proposed approach follows extractive-abstractive structure to generate summary of an event. Therefore, we review promising summarization works developed based on extractive and/or abstractive approaches.

Extractive summary of an event is the set of most relevant and informative tweets in the event. Based on desired size, summary includes number of most relevant and informative tweets. In 2010, Sharifi et al. [40] introduced a “Phrase Reinforcement (PR) Algorithm” to generate summary of one tweet length for any Twitter event. The Algorithm constructs an ordered acyclic graph with the relevant tweets as nodes and computes weight of each node from the frequency of occurrences. Maximum total weight path in the graph is generated as final summary. The work further extended to maximize the coverage of the event with more number of tweets [41] and to increase summary readability [18].

Another approach [46] constructs a weighted graph with bag of bi-grams as node and rank each node using TextRank algorithm [22] to finds top ranked bi-grams for summary generation.

Harabagiu and Hickl proposed work [13] produces a summary in 250 words rather than tweets to cover more information about an event. Important moment of any event attracts longer conversation which identifies significant and interesting tweets to produce summary [29]. Tweet text has some specific features like tweet relevance closeness to initial text, relevance regarding URL is used to measure social influence of tweets for summary generation [2]. The work by [25] explored more static social features like re-tweet information, user influence and temporal information to designate a tweet's importance and relevance to be included in summary of an event.

The work proposed in [20] collects tweets and identifies event based on the event specific activities like temporal, spatial and user behavior of that event. Then relevant tweets are ranked to generate summary. Tweet-level social information relation [14] and participant-centered social information [15] are used in recent summarization research work for generating more robust summaries. In another recent work, Chellal et al [5] formulated tweet summarization problem as optimization problem and generate summary as a subset of tweets those are more relevant, diverse and cover maximum information of the event.

Above approaches generate summaries considering all the relevant tweets as a set. But in Twitter, flow of tweets for any event varies

with the progress of time since its inception. Important moment of an event attracts more important and a good number of tweets though those tweets share similar kind of information. Due to this nature of event life in Twitter, various summarization approaches follow clustering method to summarize event. Clustering process divide an event into multiple segments and extracts most important tweets from each segment.

The work by Inouye [16] first introduced multiple cluster based event summary. But, the clustering task is challenging for Twitter due to that the cluster numbers are varies with topic. Thus clustering process highly affects summarization approaches and also needs optimization. Tweet stream clustering process follows various criteria like time window [49, 44], burstiness of tweets [53, 42, 7, 51], semantic closeness [4] and many more. Multiple approach of clustering like stream-based and semantic-based approaches are also used to clusters relevant tweets into subtopics [9] for summary generation. The work proposed in [52] used an online incremental clustering algorithm to form sub-events. Tweets in each cluster ranked based on the features like noun phrases, verbs, hashtags, URLs and numbers. Top ranked tweet from each cluster constructs the summary of the event.

An abstractive summary presents an event with key information rather than citing source tweets. With limited number of words, the abstractive summary covers more information about an event than extractive summary. The work proposed in [26] produces summary based on word graph and optimization techniques. The graph constructs with word tri-grams and weighted based on the occurrence frequencies. Rudra et al. [32] produces abstract summary for domain specific Twitter events like crisis scenario events by constructing a graph with word bigrams and word co-occurrences relation. The work [50] represents content of tweet in 5 types of speech act based on Searle's taxonomy of speech acts [37]. The algorithm extracts word and symbol based features for each tweet and labels the tweet by corresponding speech act using Support Vector Machine (SVM) classifier. Topic words and the salient words/phrases for

each major speech act type are selected by using round robin algorithm and insert into proper slots of speech act-guided templates to generate abstractive summary. Shapira et al. introduced approach [39] extracts facts of the event and expands information of the facts through alternative terms to generate summary for news event tweets.

All the previous works assume a tweet as a sentence. But in actual, a tweet includes multiple sentences in it where all of them are not contain equally important information. Earlier works eliminated redundant information by exploring semantically similar sentences only. We explored and identify partially entailed text between sentences to eliminate maximum redundant information. In this research work, we generate summaries over key sentences rather than key tweets and eliminate maximum redundant information in sentences by exploring partially entailed information.

### 3 Proposed Summarization Approach

Our proposed approach of summarization consists with two components: (1) "Key sentence extractor", which extracts most relevant and informative sentences rather than tweets from the set of event relevant tweet stream. (2) "Abstract summarizer", which generates abstractive summary over selected sentences after removing redundant information by identifying entailed or partially entailed information in sentences. Figure 1 provides an overview of the summarization approach.

The components of our summarization approach are described below.

#### 3.1 Key Sentence Extractor

Sentence identification in tweets is a very difficult and ambiguous task due to the grammatical structure of the content and inappropriate practice of punctuation. Thus, key sentence extractor component first ranks event relevant tweets based on relevance and informativeness. Next, top ranked tweets are converted into possible sentences. Finally, most relevant and informative sentences are extracted as key sentences. In the

following subsections, we elaborate the procedure of key sentence extraction.

### 3.1.1 Ranking of Tweets

A Twitter event  $E$  includes sequence of relevant tweets  $(t_1, t_2, t_3, \dots, t_n)$ . However, all the tweets are not equally important or even necessary to understand or know the event precisely. Every tweet has its own credibility to express event related information. Based on the credibility and significance, learning to rank algorithm distinguish each tweet and rank accordingly to identify most relevant and informative tweets. We follow a pair wise ranking approach in this work to list down most informative and relevant tweets for an event. We formulate the ranking task as classification problem and assign precedence to one tweet over other. For each tweet  $t_i \in (t_1, t_2, t_3, \dots, t_n)$ , algorithm assigns a rank  $y_m \in 1, \dots, n$  for a given Twitter event. Learning to rank algorithm trains a function  $h$  to measure the credibility of a tweet  $h(t)$  so that for any pair of tweets  $(t_i, y_i)$  and  $(t_j, y_j)$ ,

$$h(t_i) > h(t_j) \Leftrightarrow y_i > y_j. \quad (1)$$

We ranked tweets of an event based on the relevance and informativeness of tweet content. Relevance of a tweet to an event is the content's meaning closeness to the event and informativeness is the content's information richness to understand the event. Different features to determine the most important tweet of an event have been studied and employed in earlier works [3, 12, 11, 48]. Our ranking task draw some of the important features like length, unique words, Hashtag, Re-tweet, URL, User mention and user account features like follower, Like, List score from those previous promising tweet ranking approaches. In addition of these standard features, we introduce following 4 new features to measure relevance and informativeness of a tweet.

1. **Distribution of event keywords:** Every event has a list of words while few of them occur most frequently and very important to the event. Distribution of those most frequent words in a tweet represents its relevance to the event more closely.

2. **Semantically equivalent tweet count:** Semantically similar content can be shared through different expressions without re-tweeting a tweet. So, number of semantically equivalent tweets of a tweet is used to measure the popularity of the content.
3. **Ratio of unique words:** More number of unique words than Twitter specific words represents more rich information. This feature represents ratio of unique words with Twitter specific words in a tweet.
4. **Presence of event top hashtag:** Most frequent hashtag of an event is the most relevant hashtag. So, presence of event's top hashtag in a tweet make it more relevant to the event.

Based on above features and RankSVM algorithm proposed by Joachims [17], we propose a ranking model to rank event relevant tweets. In order to extract key sentences, we select top ranked 20 tweets for each event where number of relevant tweets are always more than 20 tweets.

### 3.1.2 Tweet to Sentence Transformation

Sentences are basic units for any summarization tool. In traditional text like news articles, sentences are easily identified due to the proper use of punctuations. But, in social media platform like Twitter, sentence boundary detection (SBD) is a difficult task. Punctuations are applied in creative and non-standard way which creates several challenges for SBD in tweets. Often a tweet includes multiple sentences in it. Using the SBD tool for social media text, developed in the work [36], we transformed each ranked tweet into possible sentences. The tool has transformed most of the tweets into possible sentences successfully. For example,

**Tweet:** Here we go #Palamedu #Alanganallur Jallikattu here ON .. We tried our best to make big with the help of @iamsridhu #Studiocr ...

**Sentence 01:** Here we go #Palamedu #Alanganallur Jallikattu here ON ..

**Sentence 02:** *We tried our best to make big with the help of @iamsridhu #Studiocr ...*

The accuracy of the SBD tool for our corpus is 84.9%. Few of the tweets where use of punctuation is ambiguous due to emphasis for multiple sentences, tool could not transformed them into possible sentences. Those tweet transformation is done through manual process.

Tweet transformation sometimes produces sentences which are not meaningful or contains only Twitter specific keywords. Thus, we follow one processing step to filter out such sentences satisfying one of the following conditions as they do not hold any key information about the event: (1) Sentence with word length less than 3, (2) Sentence containing only Hashtags, user mentions and URLs, (3) Contains only topic words.

After filtering process, we are left with selected key sentences which are most relevant and informative. Detailed statistics are reported in Table 1.

**Table 1.** Statistics of the corpus

Event Count	Ranked Tweet Count	Sentence Count	Max/Min Sentence/Event
25	500	636	36/11

### 3.2 Abstract Summarizer

In this phase, we analyze key sentences to find and remove maximum redundant information from selected key sentences.

#### 3.2.1 Fusion of Partially Entailed Sentences

Event-focused sentences often include redundant information due to the fact that similar kind of information shared through different expressions [33]. An expression may include semantically similar or partially similar content of another expressions. We have explored redundant information in sentences through identifying PTE relation between them as proposed in the work [6, 35]. We also identify the boundary of the information in one sentence that is partially entailed (PE) to the content of other sentence. This PE information among sentences is combined

to merge sentences into single sentence. In the following subsections we explain PTE identification task and the detection of PE text to merge sentences.

PTE [6, 35] is a bidirectional relationship between two expressions where one expression is entailed or partially entailed from another expression. PTE extends the concept of Textual Entailment (TE). With preservation of formal entailment definition, different categories of PTE are defined by breaking down both the expressions. This categories are PTE-I, PTE-II, PTE-III and PTE-IV. First 3 categories extends true entailment relation and last category is same as negative entailment relation. In this current work, we have generated a PTE recognition model for English tweets following the approach proposed in the work [35].

**Identification of PTE:** Our entailment decision criterion is based on various similarity scores calculated on pair of sentence. Each pair of sentence ( $s_1, s_2$ ) is represented by a feature vector, where each feature is a specific similarity score represents whether  $s_1$  or segment of  $s_1$  entails  $s_2$  or segment of  $s_2$ . We have computed 8 similarity scores for each pair of sentence. The first 4 scores are computed based on the exact lexical overlap between the words like unigram, bi-gram, longest common sub-sequence and skip-gram similarity score. The other similarity scores are cosine similarity, semantic similarity, word-to-vector similarity and soft cosine similarity [43] scores. First three features have been evaluated as effective for PTE class prediction [35] and are briefly explained below. The last feature called soft cosine similarity [43] is calculated based on the similarity of word features in Vector Space Model (VSM). This measure represents closeness of two texts more precisely by ignoring variations of words in texts.

1. **Cosine Similarity:** Cosine similarity score represents the relatedness between two n-dimensional vectors. We converted two sentences into binary vectors with values either 0 or 1 and calculated similarity score

using the following equation:

$$\cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \cdot \|\vec{B}\|}. \quad (2)$$

**2. Semantic Similarity:** We modeled the semantic similarity of two sentences as a function of the semantic similarity of the constituent words in both. Similarity score is calculated by using text to text similarity measure proposed in the work [28].

**3. Word-to-vector Similarity:** To measure word-to-vector similarity of a word, we used deep-learning library Word2Vec toolkit<sup>1</sup>. Word2Vec tool returns similarity score of two words based on a pre-trained word embedding for English tweets [27]. Word level similarity score accumulated to find sentence level similarity using following equation:

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

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

Feature profile is generated for all sentence pairs from the similarity scores to train and generate PTE identification model. We have selected a subset of the Twitter Paraphrase Corpus as in [47] for training and PTE identification model generation. This dataset is more similar to our PTE identification task and have been studied extensively to predict semantic equivalence between sentences for many NLP applications including summarization. From the chosen subset of the corpus, we manually annotate 1100 pair of sentence into four types of PTE pairs to prepare training data. Manual annotation process employed two human annotators who are post graduate students and native English speaker to annotate each pair of sentence with a category as mentioned in the work [35]. Each annotator annotates every pair of sentence to a class of PTE. Based on the similarity scores of each pair and optimization of

their relative weights, we train a Support Vector Machine (SVM) with 10-fold cross validation to generate PTE identification model. This model is applied on event-focused key sentence pairs to find PTE relation. We form sentence pairs for an event by comparing one sentence with all other sentences in the same event.

**Detection of Partially Entailed Text Boundary:** After identifying PTE sentence pairs, we merge each pair ( $s_1$  and  $s_2$ ) into a single sentence based on following cases:

1. If a pair of sentence is identified as PTE-I, then one sentence should be eliminated;
2. If a pair of sentence is identified as PTE-II, then the sentence having more information in addition of the entailed or entail information will remain in the list and other one will be eliminated.
3. If a pair of sentence is identified as PTE-III, then generate an understandable informative sentence that maximally captures the content of the both sentences.
4. If a pair of sentence is identified as PTE-IV then both the sentences should remain under consideration for further process.

From the above facts, case 1 and case 4 are straightforward where one/both sentence will remain. But case 2 and 3, where segment of a sentence is entails or entailed to other sentence or segment of sentence, boundary of the entailed text is to be identified for merging sentences. To detect the boundary of PE information, each sentence is divided into fragments i.e. piece of information. A piece of information means collection of consecutive words grouped together. A given sentence (s) can be represented as a set of its fragments like:

$S_i = F_j, F_{j+1}, F_{j+2}, \dots, F_m$ , where  $i, j = 1$  to any finite integer.

Using Ritter's Twitter tool [30, 31], we identify phrases in a sentence and consider a phrase as a fragment. In following examples we cited 2 pairs of sentences for PTE-II and PTE-III category respectively.

*PTE-II:*

<sup>1</sup><http://deeplearning4j.org/word2vec.html>

**Sentence 01:** [Its #Jallikattu day]/NP [at]/PP [the world famous #Alanganallur]/NP [.] /O

**Sentence 02:** [World Famous #Alanganallur]/NP [#Jallikattu]/HT [back]/NP [after 3 years]/NP [.] /O

PTE-III:

**Sentence 01:** [Japan]/NP [to invest]/VP [\$35 billion]/NP [in]/PP [India]/NP [over]/PP [the next 5 years]/NP

**Sentence 02:** [Investment worth]/NP [\$35 billion]/NP [to finance]/VP [infrastructure projects]/NP [and]/O smart cities]/NP [in]/PP [India]/NP [in]/PP [the next five years]/NP

We compared each fragment of a sentence with all the fragments in other sentence to identify semantically equivalent and non-equivalent fragments in the pair. Semantically equivalent fragments are recognized based on their semantic closeness [28]. Based on the semantic equivalent fragments, two sentences are merged into one as shown in figure 2 and 3. Resultant sentence is generated from partially entailed and non-entailed information.

### 3.2.2 Summary Generation

In this subsection, we explain abstract summary generation process over selected most relevant and informative sentences. In order to cover maximum information about an event in limited number of words, we generate abstract summary following the work presented in [38]. Proposed approach uses pointer generator network with coverage technique to avoid unknown words and remove repetition of words in final summary. This approach generates words for final summary from the fixed size vocabulary or source text and thus resolve unknown words generation problem in the result summary. Due to the similar attention to a particular piece of text, sometimes repetitions of words occur in summary. Coverage technique in the proposed work uses attention distribution to track attention of a word it has received at any point. The technique applies a loss term to penalize next attention to the same word and thus resolve word repetition issue. Thus, we have chosen pointer-generator network with coverage

technique for our abstract summary generation task.

## 4 Experiment Setup and Result

In this section, we present our experimental setup for assessing the performance of our proposed summarization approach. We describe our dataset, experiment setup, evaluation metrics and comparative performance.

### 4.1 Dataset

We have collected English tweets for 25 trending events using Twitter4j<sup>2</sup> during the period from January to October, 2017. Events are current happenings like natural disaster, politics, sports, entertainment and technology. In order to extract relevant tweets, we used commonly known hashtags and combination of keywords associated with those events. We tokenize each tweet using CMU tokenizer [10]. Our queries return English as well as Non-English tweets containing query hashtags and keywords. So, at initial level we cleaned all the obtained tweets by filtering out Non-English, spam and short tweets with less than 3 words. Statistics of this dataset is shown in Table 2.

**Table 2.** Statistics of the corpus

Event Nos	Tweet Nos	Filtered Tweet	Token nos
25	64,436	14,135	2,24,934

For our experiment and evaluation work, we generate gold summaries for each event consists with 100 tokens. This summary generation task is carried out by three PG students who are native English speaker. We have supplied each event relevant tweets and brief information about each event to each annotator and asked to write abstract summaries of 100 token for each event. Since, generation of gold summary for any event is a difficult task due to the diverse understanding of the event, we obtain Inter Annotator Agreement (IAA) scores as ROUGE values to evaluate the quality of

<sup>2</sup><https://github.com/Twitter4J/Twitter4J>

Sentence 01:

Phrase	Its #Jallikattu day	at	<i>the world famous #Alanganallur</i>
Fragments	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>

Sentence 02:

Phrase	<i>World Famous #Alanganallur</i>	#Jallikattu	back after 3 years
Fragments	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>

**Resultant sentence:** *Its #Jallikattu day at the world famous #Alanganallur back after 3 years*

**Fig. 2.** Fusion of PTE-II category sentence pair

Sentence 01:

Phrase	Japan	to invest	<i>\$35 billion</i>	in	India	over	<i>the next 5 years</i>
Fragments	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	F <sub>5</sub>	F <sub>6</sub>	F <sub>7</sub>

Sentence 02:

Phrase	Investment worth	<i>\$35 billion</i>	to finance	infrastructure projects	and	smart cities	in	India	In	<i>the next five years</i>
Fragments	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	F <sub>5</sub>	F <sub>6</sub>	F <sub>7</sub>	F <sub>8</sub>	F <sub>9</sub>	F <sub>10</sub>

**Resultant sentence:** *Japan to invest \$35 billion to finance infrastructure projects and smart cities in India over the next 5 years*

**Fig. 3.** Fusion of PTE-III category sentence pair

generated summary. Average F scores of ROUGE metrics for all the summaries are shown in Table 3.

## 4.2 Experimental Setting

Sentences include Twitter specific tokens like URL, Hashtag, composite word, abbreviation and Usermention etc. In order to normalization of tokens, we used social media text processing tool<sup>3</sup> developed based on the work [1] for word segmentation (for splitting hashtags) and spell correction. All the processed sentences for an event is given to pointer-generator summarization

<sup>3</sup><https://github.com/cbaziotis/ekphrasis>

tool<sup>4</sup> for summary generation. We have used a pre-trained model<sup>5</sup> having 256 dimensional hidden states, 128 dimensional word embedding, vocabulary of 50k word size and 223000 training iterations in this work for summarization tool. We set input article token limit to 600 tokens and summary token limit to 100 tokens.

## 4.3 Evaluation Metrics and Result

To evaluate the performance of our proposed approach, we used ROUGE metrics which represent the quality of machine generated

<sup>4</sup><https://github.com/abisee/pointer-generator>

<sup>5</sup><https://github.com/abisee/pointer-generator>

**Table 3.** IAA scores (ROUGE) of human generated summaries

Event	Average F-Score			
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-SU4
25	0.63	0.56	0.55	0.57

summaries comparing with human generated summaries [19]. ROUGE evaluation tool<sup>6</sup> developed based on the work [8] is used to measure ROUGE-1, ROUGE-2, ROUGE-L and ROUGE-SU4 F scores. During comparison, we considered stop words to reduce the impact of high overlap.

We conduct two fold experiments to evaluate our approach. In one fold, we report the performance of our approach (A1) in Table 4, comparing with human generated summaries. We also compare our approach with another abstract summarization tool (A2)<sup>7</sup> implemented based on sequence to sequence model with attention as proposed in [24]. Detail comparative result is reported in Table 4. The result shows that our approach outperforms over compared approach significantly.

In another fold, we compared our approach generated summaries (A1) with the summaries (A3) generated from the top ranked 20 tweets without SBD and filtering out redundant information, compared with summaries (A4) generated from the top ranked 20 tweets only without any pre-processing or normalization. The performance of comparisons are reported in Table 5. Result shows that SBD phase and PTE phase improves the quality of summaries.

Performance analysis of all the experimented approaches are also shown in figure 4. From the figure it is observed that, abstract summarization from tweets without any kind of pre-processing returns poor quality summary. This may be due to the nature of social media text and presence of Twitter specific tokens. Detailed analysis of the result is presented in next section.

<sup>6</sup><https://github.com/kavgan/ROUGE-2.0>

<sup>7</sup><https://github.com/zwc12/Summarization>

## 5 Discussion

After analyzing results, we observe that the score of ROUGE-1 is maximum and ROUGE-2 is least among all the evaluation metrics for our proposed approach as well as other experimented approaches. This may be due to the choice of word combination or word merging in the system and human generated summaries. Bi-gram overlap score sometimes decreases due to inappropriate bi-gram or order of words in sentence. ROUGE-SU4 score is also close to ROUGE-2 due to the similar kind of reason where ROUGE-SU4 measure is based on Skip-bigram and uni-gram based co-occurrence statistics.

We also observe that summaries generated from top ranked tweets without SBD and PTE include incomplete sentence or sentence with improper punctuation rather complete readable sentence. In some cases, partial redundant information is also present in the summaries. For example,

**Summary using SBD and PTE:** modi means master of developing india modi. 100 days are you satisfied with the performance so far of modi govt . the downfall of the movement ( of govt ) remains , at best , unclear .

**Summary without SBD and PTE:** modi 100 days .. modi government . modi will be successful pm & lead the nation modi. modi 100 days the direction of the movement ( of govt ) .

After through manual verification of newly generated summaries, we observe following errors:

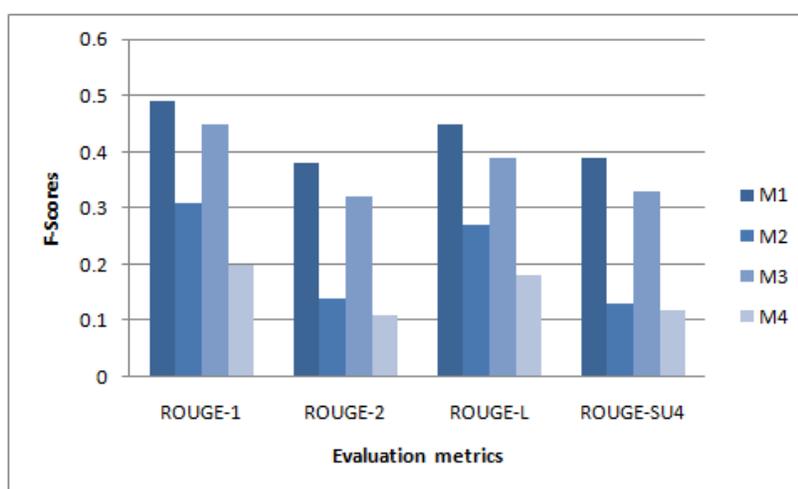
1. Our approach generated abstract summaries from extracted most relevant and informative

**Table 4.** Comparative result of proposed approach

Approach	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-SU4
A1	<b>0.49</b>	0.38	0.45	0.39
A2	0.31	0.14	0.27	0.13

**Table 5.** Performance of proposed approach on different set of data

Approach	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-SU4
A1	<b>0.49</b>	0.38	0.45	0.39
A3	0.45	0.32	0.39	0.33
A4	0.20	0.11	0.18	0.12

**Fig. 4.** Performance analysis of compared approaches

sentences. Thus, summary sentences are mostly close to the source sentence style. However, human-written summaries are mostly composed by understanding the original sentences using different vocabulary and structure. For example:

**Human generated:** *there was an earthquake in northern california of magnitude 6.0 .*

**System generated:** *i learned that there was one earthquake in northern california of magnitude 6.0 .*

- Twitter text often contain grammar, self created acronym and dictation errors. Dealing

with such texts remains a challenge for our approach. For example:

**Source sentence 01:** *usa + europe + asia must help africa , liberia cannot cope alone with ebola outbreak*

**Source sentence 02:** *this is the 8 y / o girl killed by israel in gaza this morning .*

## 6 Conclusion

Social media like Twitter is a great source of information for any happening events nowadays. Recent trend of information dissemination in Twitter makes this platform more and more popular than any news media. The main reason behind that is

fast and diverse information sharing. In this paper, we proposed an approach to summarize any event from Twitter by extracting sentences and exploring partially entailed information in sentences to avoid maximum amount redundant information. Through experimental result we showed that our proposed approach can achieve comparable result in line with the earlier research work on Twitter event summarization. Our abstractive summary cover more information within limited words and able to give a quick overview of the event.

In recent days, social media platforms are also suffering from fake information as well as information overloading problem for any happening event. To validate a tweet's trustability is a big research issue. Summary of any event must not include any information about the event which is fake or least trustable. This type of summary will create unnecessary misconception about an event. In light of the proposed approach, future scope of the work is to measure trustworthiness of tweet content to be included in the summary.

## References

1. **Baziotis, C., Pelekis, N., & Doukeridis, C. (2017).** DataStories at SemEval-2017 task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, Vancouver, Canada, 747–754.
2. **Belkaroui, R. & Faiz, R. (2017).** Conversational based method for tweet contextualization. *Vietnam Journal of Computer Science*, 1–10. ISSN 2196-8896. doi:10.1007/s40595-016-0092-y.
3. **Castillo, C., Mendoza, M., & Poblete, B. (2011).** Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*. ACM, 675–684.
4. **Chakrabarti, D. & Punera, K. (2011).** Event summarization using tweets. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*. 66–73.
5. **Chellal, A., Boughanem, M., & Dousset, B. (2016).** Multi-criterion real time tweet summarization based upon adaptive threshold. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. 264–271. doi:10.1109/WI.2016.0045.
6. **Das, A. & Rudrapal, D. (2014).** Exploring the partial textual entailment problem for bengali news texts. *Research in Computing Science*, 86, 43–52.
7. **Duan, Y., Chen, Z., Wei, F., Zhou, M., & Shum, H. (2012).** Twitter topic summarization by ranking tweets using social influence and content quality. In *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, 8-15 December 2012, Mumbai, India*. 763–780.
8. **Ganesan, K. (2018).** Rouge 2.0: Updated and improved measures for evaluation of summarization tasks. *arXiv preprint arXiv:1803.01937*.
9. **Gao, D., Li, W., & Zhang, R. (2013).** Sequential summarization: A new application for timely updated twitter trending topics. In *ACL (2)*. The Association for Computer Linguistics. ISBN 978-1-937284-51-0, 567–571.
10. **Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., & Smith, N. A. (2011).** Part-of-speech tagging for twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, HLT '11. Association for Computational Linguistics, Stroudsburg, PA, USA. ISBN 978-1-932432-88-6, 42–47.
11. **Gupta, A., Kumaraguru, P., Castillo, C., & Meier, P. (2014).** Tweetcred: Real-time credibility assessment of content on twitter. In *International Conference on Social Informatics*. Springer, 228–243.
12. **Gupta, M., Zhao, P., & Han, J. (2012).** Evaluating event credibility on twitter. In *Proceedings of the 2012 SIAM International Conference on Data Mining*. SIAM, 153–164.
13. **Harabagiu, S. M. & Hickl, A. (2011).** Relevance modeling for microblog summarization. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*. 514–517.
14. **He, R. & Duan, X. (2018).** Twitter summarization based on social network and sparse reconstruction. In *The Thirty-Second AAAI Conference on Artificial Intelligence*. 5787–5794.

15. **Huang, Y., Shen, C., & Li, T. (2018).** Event summarization for sports games using twitter streams. *World Wide Web*, 21(3), 609–627.
16. **Inouye, D. (2010).** Multiple post microblog summarization. In *Research Final Rep.* 34–40.
17. **Joachims, T. (2006).** Training linear svms in linear time. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 217–226.
18. **Judd, J. & Kalita, J. (2013).** Better twitter summaries? In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*. 445–449.
19. **Lin, C.-Y. (2004).** Rouge: A package for automatic evaluation of summaries. *Text Summarization Branches Out*.
20. **Mallela, D., Ahlers, D., & Pera, M. S. (2017).** Mining twitter features for event summarization and rating. In *Proceedings of the International Conference on Web Intelligence*. ACM, 615–622.
21. **Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., & Miller, R. C. (2011).** Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 227–236.
22. **Mihalcea, R. & Tarau, P. (2004).** TextRank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Barcelona, Spain, 404–411.
23. **Miranda-Jimenez, S., Gelbukh, A., & Sidorov, G. (2013).** Summarizing conceptual graphs for automatic summarization task. In *Conceptual Structures for STEM Research and Education*. Springer Berlin Heidelberg, Berlin, Heidelberg, 245–253. doi:10.1007/978-3-642-35786-2\_18.
24. **Nallapati, R., Xiang, B., & Zhou, B. (2016).** Sequence-to-sequence rnns for text summarization. *CoRR*, abs/1602.06023.
25. **Nasser Alsaedi, O. R., Pete Burnap (2016).** Automatic summarization of real world events using twitter. In *The Tenth International AAAI Conference on Web and Social Media, Cologne, Germany, from 17-20 May 2016*. 511–514.
26. **Olariu, A. (2014).** Efficient online summarization of microblogging streams. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2014, April 26-30, 2014, Gothenburg, Sweden*. 236–240.
27. **Pennington, J., Socher, R., & Manning, C. D. (2014).** Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543.
28. **Pirró, G. & Euzenat, J. (2010).** A feature and information theoretic framework for semantic similarity and relatedness. In *9th International Semantic Web Conference (ISWC2010)*. Springer, 615–630.
29. **Qu, Q., Liu, S., Zhu, F., & Jensen, C. S. (2016).** Efficient online summarization of large-scale dynamic networks. *IEEE Transactions on Knowledge and Data Engineering*, 28(12), 3231–3245. ISSN 1041-4347. doi:10.1109/TKDE.2016.2601611.
30. **Ritter, A., Clark, S., Mausam, & Etzioni, O. (2011).** Named entity recognition in tweets: An experimental study. In *EMNLP, EMNLP '11*. Association for Computational Linguistics. ISBN 978-1-937284-11-4, 1524–1534.
31. **Ritter, A., Mausam, Etzioni, O., & Clark, S. (2012).** Open domain event extraction from twitter. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12*. ACM, New York, NY, USA. ISBN 978-1-4503-1462-6, 1104–1112. doi:10.1145/2339530.2339704.
32. **Rudra, K., Banerjee, S., Ganguly, N., Goyal, P., Imran, M., & Mitra, P. (2016).** Summarizing situational tweets in crisis scenario. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media, HT '16*. ACM, New York, NY, USA. ISBN 978-1-4503-4247-6, 137–147. doi:10.1145/2914586.2914600.
33. **Rudrapal, D. & Das, A. (2017).** Measuring the limit of semantic divergence for english tweets. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*. INCOMA Ltd., 618–624. doi:10.26615/978-954-452-049-6\_080.
34. **Rudrapal, D., Das, A., & Bhattacharya, B. (2018).** A survey on automatic twitter event summarization. *JIPS*, 14, 79–100.

35. Rudrapal, D., Das, A., & Bhattacharya, B. (2019). Recognition of partial textual entailment for Indian social media text. *Computación y Sistemas*, 23(1).
36. Rudrapal, D., Jamatia, A., Chakma, K., Das, A., & Gambäck, B. (2015). Sentence boundary detection for social media text. In *Proceedings of the 12th International Conference on Natural Language Processing*. 254–260.
37. Searle, J. R. (1975). Indirect speech acts. In Cole, P. & Morgan, J., editors, *Syntax and Semantics 3: Speech Acts*. Academic Press, New York, 59–82.
38. See, A., Liu, P. J., & Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*.
39. Shapira, O., Ronen, H., Adler, M., Amsterdamer, Y., Bar-Ilan, J., & Dagan, I. (2017). Interactive abstractive summarization for event news tweets. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 109–114.
40. Sharifi, B., Hutton, M., & Kalita, J. (2010). Automatic summarization of twitter topics. In *National Workshop on Design and Analysis of Algorithm*.
41. Sharifi, B., Hutton, M.-A., & Kalita, J. (2010). Summarizing microblogs automatically. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10. Association for Computational Linguistics, Stroudsburg, PA, USA. ISBN 1-932432-65-5, 685–688.
42. Shen, C., Liu, F., Weng, F., & Li, T. (2013). A participant-based approach for event summarization using twitter streams. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*. 1152–1162.
43. 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*, 18(3), 491–504.
44. Wang, Z., Shou, L., Chen, K., Chen, G., & Mehrotra, S. (2015). On summarization and timeline generation for evolutionary tweet streams. *IEEE Transactions on Knowledge and Data Engineering*, 27(5), 1301–1315. ISSN 1041-4347. doi:10.1109/TKDE.2014.2345379.
45. Weng, J. & Lee, B.-S. (2011). Event detection in twitter. *ICWSM*, 11, 401–408.
46. Wu, Y., Zhang, H., Xu, B., Hao, H., & Liu, C. (2015). Tr-lda: A cascaded key-bigram extractor for microblog summarization. *International Journal of Machine Learning and Computing*, 5(3), 172–178.
47. Xu, W., Ritter, A., Callison-Burch, C., Dolan, W. B., & Ji, Y. (2014). Extracting lexically divergent paraphrases from twitter. *Transactions of the Association for Computational Linguistics*, 2, 435–448.
48. Yang, F., Liu, Y., Yu, X., & Yang, M. (2012). Automatic detection of rumor on sina weibo. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*. ACM, 13.
49. Yang, X., Ghoting, A., Ruan, Y., & Parthasarathy, S. (2012). A framework for summarizing and analyzing twitter feeds. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12. ACM, New York, NY, USA. ISBN 978-1-4503-1462-6, 370–378. doi:10.1145/2339530.2339591.
50. Zhang, R., Li, W., Gao, D., & You, O. (2013). Automatic twitter topic summarization with speech acts. *IEEE Trans. Audio, Speech & Language Processing*, 21(3), 649–658. doi:10.1109/TASL.2012.2229984.
51. Zhao, W. X., Wen, J., & Li, X. (2016). Generating timeline summaries with social media attention. *Frontiers of Computer Science*, 10(4), 702–716. doi:10.1007/s11704-015-5145-3.
52. Zhou, Y., Kanhabua, N., & Cristea, A. I. (2016). Real-time timeline summarisation for high-impact events in twitter. In *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016)*. 1158–1166. doi:10.3233/978-1-61499-672-9-1158.
53. Zubiaga, A., Spina, D., Amigó, E., & Gonzalo, J. (2012). Towards real-time summarization of scheduled events from twitter streams. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, HT '12. ACM, New York, NY, USA. ISBN 978-1-4503-1335-3, 319–320. doi:10.1145/2309996.2310053.

Article received on 05/02/2019; accepted on 23/08/2019.  
Corresponding author is Dwijen Rudrapal.