ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Template Based Abstractive Summarization of Twitter Topic with Speech Act

Gulab R. Shaikh*1, Digambar M. Padulkar2

*1 Research Scholars, 2 Asst. Prof., Department of CSE, VPCOE Baramati,Pune, India, India

shaikhgulab786@gmail.com

Abstract

Now a days, people are using microblogging services such as Twitter, Facebook, Google+ etc. The content of such site is large number of small textual messages that is posted by millions of users, at random or in response to perceived events or situations. The top trending topics on Twitter.com each can have thousands of tweets. It is very time consuming and difficult attempts to read all the tweets under a particular Trending Topic.Automatic summarization of Twitter messages (tweets) is an urgent need for efficient processing of the tweeted information. Twitter topic summarization deals with short, dissimilar, and noisy nature of tweets. In this paper, we have used a speech act-guided summarization approach.For classification purpose we have used Bagging Ensemble approch with Naive Bayes Classifier. First we have to recognize the speech acts in tweets ,then we extract key words and phrases from the tweets. An extracted key terms are ranked and inserted into special summary templates designed for speech acts. Here, in this proposed work we have implemented Ngram selection algorithm to select top ranking keyword and phrases and then inserted into summary template.

Keywords: Abstractive Summarization, Key word/phrase extraction, Speech Act, Phrase/Word ranking, Twitter, Trending Topic.

Introductions

Recently, Twitter has become one of the most popular microblogging sites. It allows user to freely post short messages (called tweets) up to 140 characters. According to a Wall Street Journal report Twitter sends out over 200 million tweets every day. The top trending topics on Twitter.com each can have thousands of tweets or more. It is impossible for a people to get an overview of important topics on Twitter by reading all tweets . The large volume of tweets suggests summarization as the key to understand the topic in single glance. Basically, a summary that provides representative information of topics with no redundancy and well-written sentences would be preferred. In this paper, we focus on the problem of twitter topic summarization with speech act, which aims to provide a template based abstractive summary for a collection of tweets on the same or similar topics.

Tweets under a Trending Topic contain a wide variety of useful information from many perspectives about important events taking place in the world. The huge number of messages, many containing irrelevant and redundant information, quickly leads to a situation of information overload. This motivates the need for automatic summarization systems which can select a few messages for presentation to a user which containes the most important information relating to the event without redundancy and filter out irrelevant and personal information.

Twitter is highly attractive for information extraction and text mining purposes, because they offer large volumes of real-time data. The quality of messages varies significantly, however, ranging from high quality text to meaningless strings. Typos, ad hoc abbreviations, phonetic substitutions, ungrammatical structures and emoticons etc.

Social Networking is a lightweight and easy form of communication. Microblogging sites such as Facebook, Google+ and Twitter, has become ubiquitous in its use with over 4 billion mobile devices worldwide of which over 1 billion support smart services.

Twitter topic summarization is substantially different from summarization of other forms of text such as news articles, research papers, books, etc because of the following reason.

Twitter topic summarization is multi-document summarization approch. Typically, multi-document summarization deals with dozens of documents each with several hundred of words .On the other hand, the tweets under a given topic usually consists of only one or two sentences and each tweet containe only 140 characters.

Multi-document summarization consist of closely related documents, such as news reports about the same event that overlap considerably in their

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

contents. On the other hand, tweets under a topic are not related to each other.

The texts from News articals, Books, Research paper etc. are usually formal writings and have high language quality. On the other hand, the language of tweets is highly noisy, spelling and grammar mistakes. Tweets containe Typos, ad hoc abbreviations, phonetic substitutions, ungrammatical structures and emoticons etc. Due to the above charac teristics, text summarization techniques in general may not adapt well to the Twitter text.

Literature survey

Phrase Reinforcement Algorithm

For a Given a trending topic, one can find twitter post i.e tweets from Twitter.com that contain the topic phrase. Inorder to generate the summary user would have to read these posts and manually summarize their content. Instead, here author automate this process using Phrase Reinforcement Algorithm.

The main task of the Phrase Reinforcement (PR) algorithm is to find the most commonly used phrase that containe the topic phrase. Then this phrase is used as a summary. PR algorithm was inspired from following two simple observations:

(1) While describing a key idea users will often use the same word or sets of words adjacent to the topic phrase.

(2) users will often "re-tweet" the most relevant content for a trending topic.

The Phrase Reinforcement algorithm[7] begins with a starting phrase. Given the starting phrase, PR algorithm submits a query to Twitter.com for a list of posts that each contains the phrase. Once the posts are retrieved, next work is to filters the posts to remove any spam or other sources of irrelevant data . Filtering is an important step in order to focus the algorithm on the most relevant content. Finally, isolate the longest sentence from each post that contains the topic phrase. These sentences form the input into the PR algorithm.

The PR algorithm starts by building a graph that represent the common sequences of words (i.e. phrases) that occur both before and after the topic phrase. The graph is generated such that it centers about a common root node representing the topic phrase. Adjacent to the root node are chains of common sequences of words found within the input sentences.

Generating Summary of Multiple News Article

SUMMONS system[12], is the first summarization system that is able to summarize several documents at once. The goal is to produce a single summary that merges all the relevant information about that event and how reports evolved over time.

The architecture of the SUMMONS system consists on two major components:

1.Content Planner: It selects the information to include in the summary by properly combining the input templates.

2.Linguistic Generator: It selects the right words to express that information and arranges them to form English sentences.

Content planning, is performed by using specifically devised "summary operators". These consist on a set of heuristic rules that are built from corpora.

Working of the SUMMONS System:

In SUMMONS, initially every input templates are given equal importance. Pairs which overlap are then identified and one of the above operators runs on them and performs a decision, either combining them into a single template, requiring that one appears earlier in the summary than the other, providing a logical connection between them, etc. This is done until no operator can be applied. At the end, the linguistic generator gathers all the combined information and uses connective phrases and cue phrases to synthesize a summary.

Extractive Summarizer: MEAD

An Extractive Summarization approach is represented by MEAD[2], which is adapted from the open source summarization framework. MEAD decides which sentences to include in the extract by ranking them according to a set of parameters. The input to MEAD is a cluster of articles (e.g., extracted by CIDR), segmented into sentences and a value for the compression rate R. The output is a sequence of n * r sentences from the original documents presented in the same order as the input documents.

For example, if the cluster contains a total of 100 sentences (n = 100) and the value of R is 20%, the output of MEAD will contain 20 sentences. Sentences appear in the extract in the same order as the original documents are ordered chronologically. We benefit here from the time stamps associated with each document.

To compute the salience of a sentence author have used following three features:

1.Centroid value

2.Positional value

3.First-sentence overlap.

Advantages and Limitations:

Author have used a new utility-based technique, RU, for the evaluation of MEAD .Here, author found that MEAD produces summaries that are similar in quality to the ones produced by

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

humans. MEAD [2] does not have a special mechanism to deal with controversial features. It is not clear how overall controversiality of a feature can be effectively expressed with extraction, as each sentence conveys a specific and unique opinion. One could include two sentences of opposite polarity for each controversial feature. However, in several cases that we considered, this produced extremely incoherent text that did not seem to convey the gist of the overall controversiality of the feature.

Implementation detail

System Architecture:

System Architecture of Twitter Topic Summarization is shown in Fig.3.1. Twitter Topic Summarization system consist of 3 core modules:

1. Recognizing speech acts in tweets.
2. Extracting speech act-guided key words/phrases and.
3. Generating abstractive summaries.

Feature Set Design:

In the following, we have described the feature sets used for recognizing the five types of speech act i.e Statement,Comment,Question,Suggestion,Commit including word-based and symbol-based features.

1.Word-Based Features:

Cue Words And Phrases :

Some speech acts are typically signalled by some cue words or phrases, such as whether for question and could you please for suggestion.

Non-cue Words:

There are four types of non-cue words which are explained in the following.

1.Abbreviations And Acronyms:

While communicating through twitter user may use some abbreviations. Examples are GMfor Good Morning and tq for thank you. Before moving towards next step we need to restore the shortened words to their original forms.

2.Opinion Words:

The Senti-WordNet and Wilson Lexicon are used for opinion mining. While communicating through twitter People may use strong opinion words.

3.Vulgar Words:

People may use some vulgar words such as c**t and f**k.

4.Emoticons:

While twittering people use emoticons to express his expression such as O:)
and *-*.

Symbol-Based Features:

Here, there are two types of eight symbol-based features, which indicate the frequency and position of special characters and are either binary- or ternary-valued.

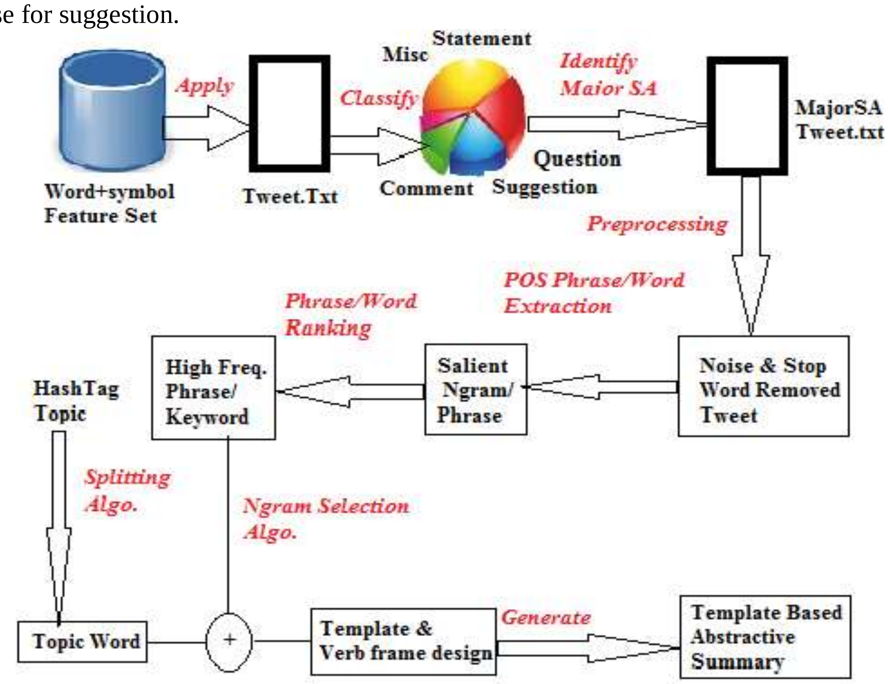

Fig.3. 1. Architecture of the Proposed System

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

Twitter-Specific Symbols:
There are three twitter specific symbols such as: #,@, and RT.
Is a hashtag marker often used in a mention of something to be stated about or commented on.
@ Is a prefix to a tweeter account, which tends to be associated with the more interpersonal speech acts questions or suggestion.
RT stands for "Retweet" and its presence, especially in the initial position, strongly indicates a statement.

Each of those symbols is associated with two features: one binary-valued fea- ture indicating whether the symbol is in the initial position of a tweet and one ternary-valued feature indicating whether the symbol does not appear (0), appears one or two times (1), or ap-pears more than two times (2).
Indicative Punctuations:
There are two Indicative punc- tuations: ? and !
? -Indicate a Question.
! -Indicate a Comment or Suggestion.
Each of them is associated with 1 ternary-valued feature indicating zero appearance (0), one or two appearances (1), or three or more appearances (2).

Speech Act-Guided Key Word/ Phrase Extraction

In this proposed system, the main purpose of Twitter speech act recognition is to sort out the tweeted content for extracting summary-worthy information. Here in this work focus on only 5 real types (Statement, Comment, Suggestion, Question, Commit) and extract key phrases and words from the tweets. For the major speech act we have extracted speech act relevant keyword from the POStagged tweets.

Noise -Resistant Phrase Extraction

To extract key words and phrases from the tweets of major speech act types, first need to compile a stopword list to filter less informative words. Then we extract key words as frequent nonstop words. Extracting the key phrases is formulated as finding frequent ngram collocations.

POS-Based Phrase/Word Patterns

Here, First need to consider that not all the extracted key words and phrases convey the most relevant information to a speech act. For example, statements are about facts, things, people, etc. and suggestions are about actions, activities, etc. Such information can be approximated by part-of-speech (POS) patterns for both words and phrases.Here we have used Stanford POStagger for tagging the each word in tweets.

A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word, such as Noun(NN), Verb(VB), Adjective(JJ),Adverb(RB) etc.
Representative POS-based regular expression patterns are listed in the following, along with illustrative examples.
The Statement-relevant word is a noun, or /N/ (e.g., school), phrase is a noun phrase, such as /Adj/ /N/ (e.g., high quality)
The Comment-relevant POS patterns are like the statement- relevant ones. But comment phrases must have at least one opinion word (e.g., bad thing).
The Suggestion-relevant word is a verb, or /V/ (e.g., hate), phrase is verb centered, such as /Adv/ /V/ (e.g., truly wish) and /V/ /N/ /N/ (e.g., sell health drugs).
The Question-relevant word is either a verb or a noun, or (/N/ /V/) (e.g., reason), phrase is either a noun phrase or a verb centered phrase, such as /Adj//N/ /N/ (e.g.,high quality drug).

Phrase/Word Ranking

First, select the most salient speech act-relevant words and phrases (ngrams) for generating a summary. In this work, salience is understood as a cumulative effect from an ngram network,i.e. a salient ngram co-occurs with other salient terms in the same tweet, which in turn boosts the salience of other ngrams it co-occurs with.

After extracting the speech act relevant keyword ,we construct a graph G for the whole tweets of a major speech act type, using all the extracted ngrams (Ng) as vertices. Two vertices N gi and N gj are linked by an edge if they co-occur in some tweet and the weight of the edge(wij) is the number of such co-occurrences.

Here G is undirected graph and use N B(N gi) to denote the neighborhood of N gi.Finally, define the graph score of N gi, GS(N gi)as:

$$GS(Ngi) = \frac{1-d}{|Ng|} + d * \sum \frac{GS(Ngi) * Wij}{\sum_{ngk \in NB(Ngj)} Wkj}$$

The graph vertices are generally unigrams from which phrases are later assembled. Phrases are more informative and less ambiguous than words (compare college life with college or life).
So count the length Ni of N gi into its salience score SS(Ngi).
SS(N gi) = GS(N gi) * Ni
and then rank all the phrases and words. Here rankings are determined by salience scores for each phrase and each word.

Twitter topic summarization

For a Twitter topic summarization, high ranking keywords and phrases extracted for its major speech act types and Twitter trending Topic name

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

itself is the building blocks of our template based abstractive summary. This generated summary is abstractive in nature as proper words/phrases are to be filled in slots of a template specially designed to accommodate speech acts and speech act verbs.

Template Design

With the topic words and the salient words/phrases for each major speech act type, generate an abstractive summary by inserting them into proper slots of speech act-guided templates.

The following figure shows template for abstractive summary.

for”<topicwords>”,people<verbframe>”<ngrams>”{,(and)<verb frame>”<ngrams>”}*.

In above figure boldfaced words and punctuations are template constants and the angle brackets(<>) enclose template slots to be filled; (and) means the word and is optional; {}*means the enclosed part can appear zero or one or more times. The topic words are derived from the topic. For a regular topic, they are a direct copy; for a hashtag topic, they are the split result of the hashtag.

The ngrams are the salient words/phrases extracted for the major speech act types. A verb frame is a verb or verb phrase specific to a particular speech act type. The speech act type and verb frames are listed in TABLE 1.

Table 1: Speech Act & Verb

Speech Act	Verb Frame
Statement	state
Comment	comment on
Suggestion	suggest
Question	ask about

Proposed implementation approch

Bagging Ensemble approach is used to solve the problem of scalability, since ro- bust and accurate classifier aggregations can be learned even if each individual classifier operates on the incomplete training data.

Twitter topic summarization is carried out with following major steps:

First, collect the twitter tweet on particular trending topic from Twitter.com using twitter4j API.

Use Bagging Ensemble with Naive Bayes classifier to classify the tweets according to speech act.

Identify the major speech act and apply Natural Language Preprocessing on tweets.

Apply Stanford POStagger and extract the speech act relevant keyword and phrases from POS tagged tweets.

Select high ranking keywords and phrases according to Graph Score of each extracted keyword.

Finally we are going to generate template based abstractive summary .

In this work, we are using Naive Bayes ensemble with bagging technique to solve the problem of scalability. We train each Naive Bayes independently over the replicated training dataset and then aggregate the trained Naive Bayes via Majority Voting scheme.

Bagging is the procedure that generates k no.of bootstrap samples from training data, and each one is taken as a subset . A bootstrap sample is generated by randomly sampling N examples from the training dataset with replacement, where N is the size of training dataset.

The algorithmic steps of constructing the ensemble Naive Bayes with bagging is shown in following algorithm:

1. procedure BAGGING()
2. Input: A training set S = (x1 , y1), . . . , (xn , yn), where xi ∈ x and yi ∈ y = (1, 2, .. . , L); K : the number of base classifiers;
3. for i ← 1 to k do
4. Generate a training set S = bootstrap sample with replacement from S;
5. Apply Naive Bayes on S ,Ck : X −−· Y ;
6. end for
7. Output the aggregated classifier based on Majority Voting approach.
8. end procedure

Proposed Algorithm:

Hashtag Splitting Algorithm:[1]

1. procedure SPLIT(Hm) H is the hashtag Contains m character
2. Split (H0) is empty;Split (H1) is H’s first character itself;
3. for i ← 2 to m do
4. for j ← 0 to i − 1 do
5. Calculate score (fj) where fj is formed by Split(Hj) and a ”word” as the remaining part of Hi,with Hj removed;
6. Choose the highest scoring fj to be Split(Hi);
7. end for
8. Output Split(Hm) ie..Split(H)
9. end fo
10. end procedure

Ngram Selection Algorithm:[1]

1. procedure NGRAMSELECTION()
2. repeat
3. for all SpeechAct in template do
4. Select the top-ranking N g* from all the ngrams extracted for that speech act

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

5. if N g* is a unigram the
6. Skip to the next speech act unless all longer ngrams (length ≥ 2) for all speech acts have been selected;
7. end if
8. if N g* is not redundant and summary length permits then
9. Fill a template slot with N g*;
10. Else
11. Remove N g*;
12. end if
13. end for
14. until summary length is reached;
15. end procedure

Result

Dataset

We have collected tweets on following latest Trending topic from Twitter.com using twitter4j API on various News and Entity topic. Each tweet is manually labelled according to speech act i.e Statement,Suggestion,Question,Comment and Commit. TABLE 2 shows the details of Experimental Twitter Dataset.

Table 2:Details of Experimental Twitter Dataset.

TrendingTopic	Tweets
Narendra Modi	1836
Barack Obama	1200
Rahul Gandhi	1024
Aakash And Iphone	2000
India	786

.

Output

Output Statistic On Training Dataset

In Figure 6.1 we have shown the output statistic such as Precision and Recall value on each speech act i.e Statement, Suggestion, Question, Comment, Commit.

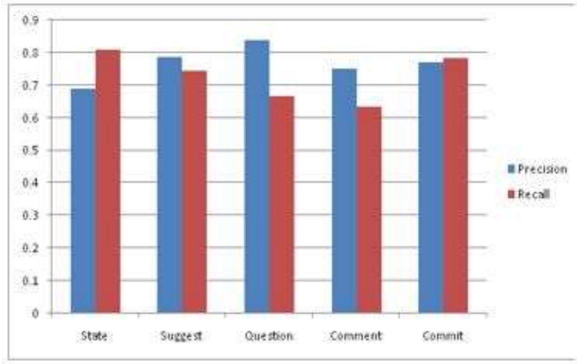

Figure 6.1:Precision and Recall.

In figure 6.2 we have shown classifier accuracy in percentage with respect to number of iterations. Above chart shows the correctly classified instance and incorrectly classified instances with no.of iterations

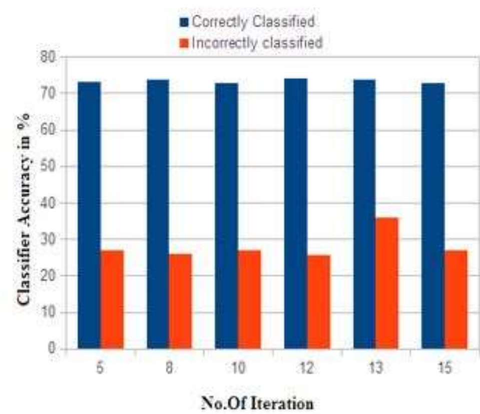

Figure 6.2:Classifier Accuracy with No.Of Iteration.

Part Of Speech Tagging

Figure 6.3 shows the output snapshot of Stanford POStagger.A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word, such as Noun(NN), Verb(VB), Adjective(JJ),Adverb(RB) etc.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

```
run:
Reading POS tagger model from TaggerFile/english-left3words-distsim.tagger ... done [2.6 sec].
Toon_NNP :_: See_VB SECULAR_NN http://t.co/QFqPtK8E_NN
The_DT day_NN similar_JJ officer_NN probing_VBG Gujrat_NNP riots_NNS Ur_NNP Narendra_NNP
Modi_NNP ._.
Local_JJ Gujrat_NNP I_PRP
Trust_NNP NDTV_NNP bring_VB Gujrat_NNP 2002_CD article_NN Narendra_NNP Bhopal_NNP
rally_NN article_NN http://t.co/r4VPI2bPbo_NN
-_: riot_NN 13_CD Years_NNS ,_, Responsible_JJ person-Narendra_NN Modi_NNP In_IN UP_NN
36_CD Riots_NNS 1_CD Year.Responsible_NN person-SOCIAL_NN Media_NNP
9_CD Narendra_NNP modi_NN roaming_NN country_NN cities_NNS gujrat_VBP surat_NN
drowning_VBG flod_NN ._.
UPA_NNP govt_NN applying_VBG means_NNS discredit_VBP succeed_VB evil_NN
```

Figure 6.3:POStagged Tweets.

Speech Act Relevant keyword Extraction and Graph Score

Figure 6.4 shows the speech act relevant keyoword extraction snapshot. Depending on the majority speech act we are going to extract the phrases and keyword from Part Of speech(POS)Tagged tweets. Figure 6.5 shows the Graph Score of each extracted keyword.

```
run:

I found the text: savVy_JJ leader_NN narendra_NN

I found the text: similar_JJ officer_NN

I found the text: Local_JJ Gujrat_NNP

I found the text: due_JJ arrival_NN
```

Figure6.4:Statement Relevant Keyword

```
GS[savVy][1] =1.7166666666666666
GS[leader][0] =0.8666666666666667
GS[narendra][3]=0.8666666666666667
GS[similar][5] =0.8666666666666667
GS[officer][7] =0.8666666666666667
GS[Local][2] =0.016666666666666667
GS[Gujrat][6] =0.016666666666666667
GS[due][4] =0.016666666666666667
GS[arrival][8] =0.016666666666666667
```

Figure6.5:Graph Score

Abstractive Summary Template

Following output snapshot shows the template of abstractive summary. With the topic words and the salient words/phrases for each major speech act type, generate an abstractive summary by inserting them into proper slots of speech act-guided templates.

```
FOR Narendra Modi PEOPLE state

savVy leader narendra similar officer Local Gujrat due arriva
```

Figure 6.6:Template Based Abstractive Summary.

Conclusion and future work

In this proposed work, speech act-guided summarization approch is used to generate summary of twitter trending topic. With the recognized speech acts, next step is to extract key words and phrases from tweets to generate abstractive summaries. The extracted key terms are then ranked and inserted into special summary templates designed for speech acts, by using ngram selection algorithm.This system is designed to accommodate the numerous, short, dissimilar, and noisy nature of the tweets. This proposed approach makes a good work contribution to the summarization community.

Twitter Topic Summarization system used on Twitter trending topics.There are two types of twitter trending topic, Regular and Hashtag topic. Template based abstractive summarization method generate summary that are significantly more explanatory, informative,short and readable than extractive summaries.

In future, this proposed system can ex- tended to electronic media such as email and discussion forum in order to study the behavior of email or message senders. In future, results of Twitter Topic Summarization can be improved by experimenting with different classifiers such as descision tree and

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

with ensemble learning such as Bagging,Boosting etc.

References

1. Renxian Zhang, Wenjie Li, Dehong Gao, and You Ouyang, Automatic Twitter Topic Summarization With Speech Acts,IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 21, NO. 3, MARCH 2013
2. D. Radev, H. Jing, M. Sty, and D. Tam, Centroid-based summa- rization of multiple documents, Inf. Process. Manage., vol. 40, pp. 919938, 2004
3. Dehong Gao, Wenjie Li and Renxian Zhang.Sequential Summarization: A New Application for Timely Updated Twitter Trending Topics.,51 ACL Sofia, Bulgaria, August 4-9, 201
4. R. Zhang, D. Gao, and W. Li, What are tweeters doing: Recognizing speech acts in twitter, in Proc. AAAI-11 Workshop Analyzing Micro- text, 2011
5. G. Carenini and J. C. K. Cheung, "Extractive vs. NLG-based abstractive summarization of evaluative text: The effect of corpus controversiality," in Proc. ACL-08, 2008, pp. 33–40.
6. B. Sharifi, M.-A. Hutton, and J. Kalita, Summarizing mi- croblogs au- tomatically, in Proc. HLT/NAACL-10, 2010.
7. B. Sharifi, M.-A. Hutton, and J. Kalita, Experiments in microblog summarization, in Proc. IEEE 2nd Int. Conf. Social Comput., 2010.
8. B. Sharifi, M.-A. Hutton, and J. Kalita, Automatic Sum- marization of Twitter Topics, in National Workshop on Design and Analysis of Algorithm, Tezpur, India, 2010.
9. G. Erkan and D. Radev, LexRank: Graph-based Centrality as Salience in Text Summarization, J. Artif. Intell. Res., vol. 22, pp. 457479, 2004.
10. R. Mihalcea and P. Tarau, TextRank: Bringing order into texts, in Proc. EMNLP-04, 2004, pp. 404411.
11. X. Wan and J. Yang, "Multi-document summarization using cluster based link analysis," in Proc. SIGIR-08, 2008, pp. 299–306.
12. Kathleen McKeown and Dragomir R. Radev (1995),Generating Summaries of Multiple News Articles. In Proceeding of 18 annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.74-82,Washington
13. G. Erkan and D. Radev, "LexRank: Graph-based Centrality as Salience in Text Summarization," J. Artif. Intell. Res., vol. 22, pp. 457–479, 2004.

Author Biblography

Prof. Digambar M. Padulkar
is working as Assistant Professor, Department of Computer Engineering ,Vidya Pratishthans College of Engineering, Baramati, Pune, Maharashtra, India. He has 10 years of experience in teaching and research. His research areas include Uncertain Data Mining, Application of Data Mining in Business and Intelligent predictions. He has publications at international journals/conferences.
Currently He is working on projects on ensemble approaches. He is life-member ISTE, CSI.
Email: dm.padulkar2006@gmail.com

Gulab Rajabhai Shaikh
Received the Bachelor's degree (B.E.) in IT in 2009 from VPCOE, BARAMATI. He is now pursuing ME degree, in Computer Engineering at VPCOE, BARAMATI. Gulab has published paper on " A Survey on Template Based Abstractive Summarization of Twitter Topic Using Ensemble SVM with Speech Act",in International Journal of Engineering Research and Technology.His current research interests include Data Mining, Image Processing etc.
Email: shaikhgulab786@gmail.com ,