SOCIAL MEDIA ANALYTICS FOR INDIAN RAILWAY

Pratish Bobde[1]. Rahul Vairal[2], Navnath Jadhav[3], Prasad Shinde[4]

[1]*Information Technology, Sanjivani COE, Kopargaon, bobde71@gmail.com*
[2]*Information Technology, Sanjivani COE, Kopargaon, rahulvairal777@gmail.com*
[3]*Information Technology, Sanjivani COE, Kopargaon, navnathujadhav@gmail.com*
[4]*Information Technology, Sanjivani COE, Kopargaon, prasadshinde9657@gmail.com*

Abstract

Considering wide use of twitter as the source of information, reaching an interesting tweet for a user among a bunch of tweets is challenging. Many private and/or public organizations have been reported to create and monitor targeted Twitter streams to collect and understand user's opinions about the organizations. However the complexity and hybrid nature of the tweets are always challenging for the information retrieval and natural language processing. Targeted Twitter stream is usually constructed by filtering and rending tweets with certain criteria with the help proposed framework. By dividing the tweet into number of parts "targeted tweet" is then analyzed to the understand users opinions about the organizations. There is an emerging need for early rending and classify such tweet, and then it get preserved on dual format and used for downstream application. The proposed architecture shows that, by dividing the tweet into number of parts the standard phrases are separated and stored so the topic of the tweet can be better captured in the subsequent processing of the tweet proposed system on large-scale real tweets demonstrate the efficiency and effectiveness of our framework.

Keywords—Tweet Segmentation, Information Retrieval, Named Entity Recognition.

I. INTRODUCTION

Twitter, as a new type of social media, has seen tremendous growth in recent years.It has attracted great interests from both industry and academia. Many of the private and/or public organizations can have the social links also have been reported to monitor Twitter stream to collect and understand user's opinions about the organizations. Nevertheless, due to the extremely large volume of tweets published every day, it is practically infeasible and unnecessary to listen and monitor the whole Twitter stream. Therefore, targeted Twitter streams are usually monitored instead; each such stream contains tweets that potentially satisfy some information needs of the monitoring organization. Targeted Twitter stream is usually constructed by filtering tweets with user-defined selection criteria depending on necessity of users requirement.

For example, the criterion could be a region so that user's opinions from that particular region are collected and monitored; it could also be one or more predefined keywords so that opinions about some particular events/topics/products/services can be monitored. The idea is to segment an individual tweet into a sequence of consecutive phrases, each of which appears more than chance.

II. Literature Survey

Chenliang Li, Aixin Sun, JianshuWeng, and Qi He [2015] in their paper entitled "Tweet Segmentation and its Application to Named Entity Recognition" have discussed and proposed methods for tweet segmentation Microblogging sites such as Twitter have reshaped the way people find, share, and disseminate timely information. Many organizations have been reported to create and monitor targeted Twitter streams to collect and understand users opinion. Targeted Twitter stream is usually constructed by filtering tweets with predefined selection criteria (e.g., tweets published by users from a geographical region, tweets that match one or more predefined keywords). Due to its invaluable business value of timely information from these tweets, it is imperative to understand tweets

language for a large body of downstream applications, such as "Named Entity Recognition"(NER) event detection and summarization, opinion mining, sentiment analysis and many others. Given the limited length of a tweet (i.e., 140 characters) and no restrictions on its writing styles, tweets often contain grammatical errors, misspellings, and informal abbreviations. The error-prone and short nature of tweets often make the wordlevel language models for tweets less reliable. For example, given a tweet I call her, no answer. Her phone in the bag, she dancing, there is no clue to guess its true theme by disregarding word order (i.e., bag-of-word model). The situation is further exacerbated with the limited context provided by the tweet. That is, more than one explanation for this tweet could be derived by different readers if the tweet is considered in isolation. On the other hand, despite the noisy nature of tweets, the core semantic information is well preserved in tweets in the form of named entities or semantic phrases.

III. System Analysis Proposed Architecture

In social media, data is widely spread across all over the world. Data is generated day by day in large amount of form. So this data cannot be handled manually so we have developed an application for Indian Railway to improve their performance. This Application makes easy to understand the people requirement about Railway. In this Architecture firstly input the Social Media Data then it is collected in one particular file. Then this file puts into the clustering form. The Indian Railways having specific keywords this makes sense to our system. These keyword makes the filtering of sentences. The data is collected in different forms such as blogs, posts this classify in the form of appreciation suggestion or in the form of good or bad comments. Then these comments analyze by the framework. The information is predicted by the system. This data comes into the summarized form.

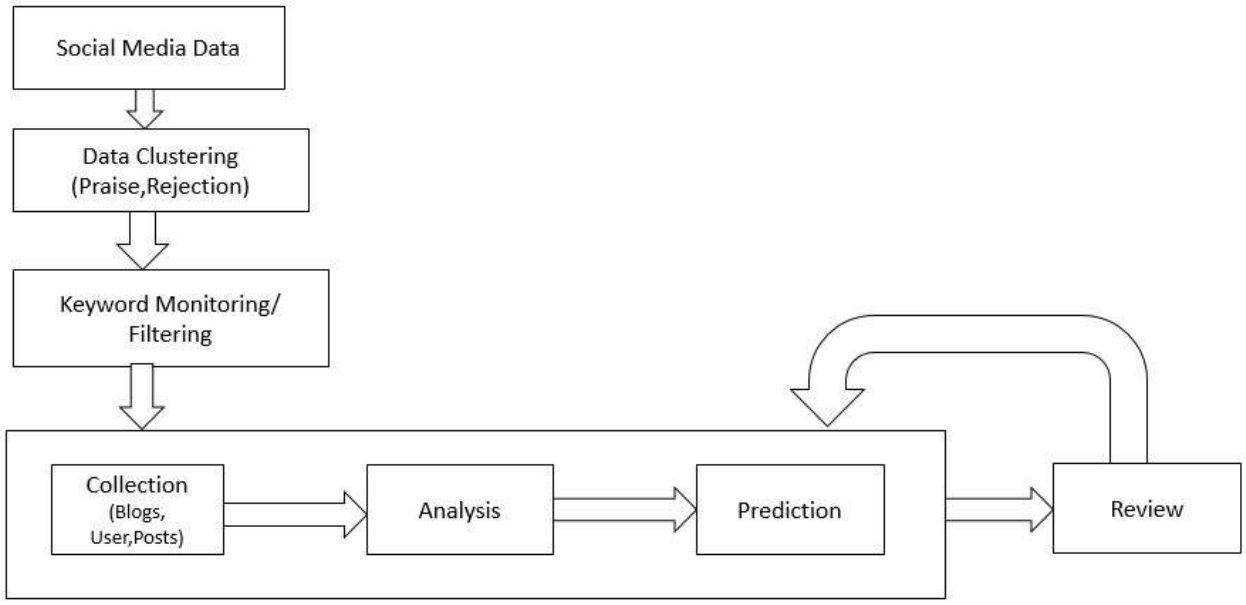

Fig. System Architecture

PROPOSED SYSTEM:

1) In system focus on the task of tweet segmentation. The goal of this task is to splita tweet into a sequence of consecutive n-grams, each of which is called a segment. A segment can be a named entity (e.g., a movie title finding

nemo), a semantically meaningful information unit (e.g., officially released), or any other types of phrases which appear more than by chance.

2) To achieve high quality tweet segmentation, propose a generic tweet segmentation framework, named HybridSeg. HybridSeg learns from both global and local contexts, and has the ability of learning from pseudo feedback.

3) Global context. Tweets are posted for information sharing and communication. The named entities and semantic phrases are well preserved in tweets.

4) Local context. Tweets are highly time-sensitive so that many emerging phrases like She Dancing cannot be found in external knowledge bases. However, considering a large number of tweets published within a short time period (e.g., a day) containing the phrase, it is not difficult to recognize She Dancing as a valid and meaningful segment. Therefore investigate two local contexts, namely local linguistic features and local collocation.

IV. CONCLUSION

To presents a prototype that supports continuous tweet stream summarization. A tweet stream clustering algorithm to compress tweets into clusters and maintains them in an online fashion. Then, it uses a Rank summarization algorithm for generating online summaries and historical summaries with arbitrary time durations. The topic evolution can be detected automatically, allowing System to produce dynamic timelines for tweet streams by using Local and Global Context. Though the proposed framework show the segment-based named entity recognition methods achieves much better accuracy than the word-based alternative. It would be interesting future work to exploit similar ideas to support further improvement to the segmentation quality by considering more local factors. The other is to explore the effectiveness of the segmentation-based representation for tasks like tweets summarization, search, hash tag recommendation, etc. Also there is significant difference in performance indicates the language mismatch problem by managing the segmentation accuracy and system performance it could be handle.

REFERENCES

[1] Tweet segmentation and its application to named entity recognition; Chenliang Li, Aixin Sun, JianshuWeng, and Qi He

[2] Using analytics and social media for monitoring and mitigation of social disasters;Horia Nicolai Teodorescu

[3] Exploring the Use of Social Media During the 2014 Flood in Malaysia;TengkuSitiAishaa, SaodahWokb, AiniMaznina A Manafc, RizalawatiIsmaild

[4] Social Media for Situational Awareness: Joint-Interagency Field Experimentation; Scott Appling, Erica Briscoe, Ann Carpenter, Leigh McCook, Gerald Scott, Tristan Allen, Raymond Buettner, Carl Oros