Trend Identification in Social Network
(Globally or Locally)

[1]Vishal D. Monpara, [2]Prof. Pradyumansinh U. Jadeja

[2]Assistant Professor
[1]Computer Engineering Department,
[1]Darshan Institute of Engg. & Tech., Rajkot,India

_____

Abstract - **As usage of IT technologies increasing day by day, the quantity of web generated data are also increasing. This leads to large data collected in to web. So, different web mining techniques can be applied to explore and analyze web data for extraction of interesting information. Twitter is popular online social networking website which provides users to post and share tweets with restriction of 140 characters. Day by day tweets and number of active users are increasing on twitter and which generates lots of web data that can be useful for the detection of trends. The content shared by user can be globally or locally, which forms the Trend in community. It is important for organization to know these trends as quickly as possible. Many methods and systems have been proposed for trend identification over social network. We have proposed a new method for detection of trend from the various tweets collected from twitter users.**

IndexTerms – **Web Mining, SNS, Trend Identification, Social Network.**
_____

I. INTRODUCTION

Web mining is increasingly becoming popular among IT professionals, academics and pupils. In the last years, almost in every meeting which has anything to do with databases, neural networks, genetic algorithms, E-commerce, or artificial intelligence has had a theme or session on Web mining. From beginning, the World Wide Web are changing rapidly, when a user starts to user web. Every day new technologies are coming for enhancement of Web and Network, Hardware cost and raw material cost of computer development decreases, also internet connections prices are going down day by day, this leads to mass use of internet by people.

The websites like social networks, blogs, forums provides a platform for sharing contents and information between different users. This leads to generation of user data which can be extremely helpful in different scenario. Our work explores the use of user generated information for the identification of trends in social network. This work narrows the subject domain to social network only; also the same working methods can be applied to any other web data domain which is likely in interactions with online social users.

II. PROBLEM STATEMENT

Trend identification in social network is useful for the prediction future events. Usage of social network is increasing day by day; contents shared and posted by users are increasing. Using this information we can predict the current and future trend which can be helpful for the researchers as well as industrialist.

III. PROPOSED SYSTEM

The data collected from social network platform is useful for detecting trend over social network. In proposed work system the trend accuracy can be higher than the existing system. Below diagram shows the proposed work architecture.
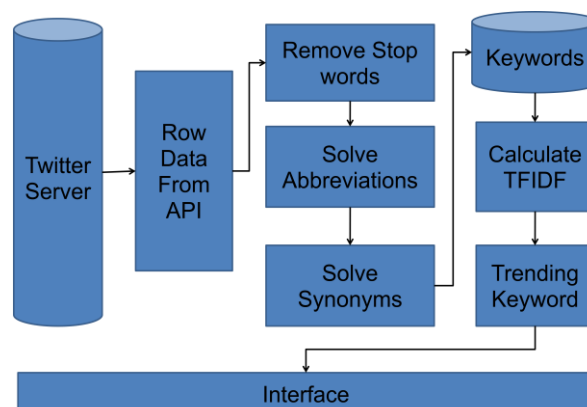

Figure 1 Proposed Work Architecture

In our proposed architecture we first collect the Data from API of the social networking website. This data is unstructured and contains the Keywords. These keywords can be useful for the detecting trend for particular time span.

The keywords are extracted from the data and frequency for all keywords are calculated for the given time period. These keywords are stored into the database.

On these keywords different operations are performed for the Trending keywords detection. These operations are like removing typing mistakes, solve abbreviations, and solve synonyms from the keywords. After removals of stop words and general keywords we can find the Trending keyword which can be the Trend for the particular social network.

IV. IMPLEMENTATION AND ANALYSIS

The proposed system will implemented with C#.Net programming language and Sql Server as a database. For reading of different tweets and followers of the user we need to access the twitter REST api. Twitter api gives access by Oauth enabled application standard, this can be achieved by creating application on twitter developer account and creating Access Token, Access Token Secret, API Key, API Secret Key.

Figure 2 Twitter Access Token

Getting the Tweets

Using twitter api we got the tweets of a particular user by providing the UserID and since_id parameter to the api query string. Below is the screen shot which demonstrate the query construction and execution using api.

Removing Mentioned User in Tweet

After collection of tweets we have to remove the mentioned user from the tweet for the better trend identification. User can mention the user in their tweets by "@" followed by their tweeter username

Removing URLs from Tweet

User can include the third party urls in their tweets also. Twitter has inbuilt functionality which converts any url mentioned into the tweets into the http://t.co/ format. These urls can be removed from the collection of tweets.

Removing Hash Tag

Hash tags are used for identifying particular information from the tweets. These hash tag can be defined into tweets by "#". We can remove these hash tags from tweets.
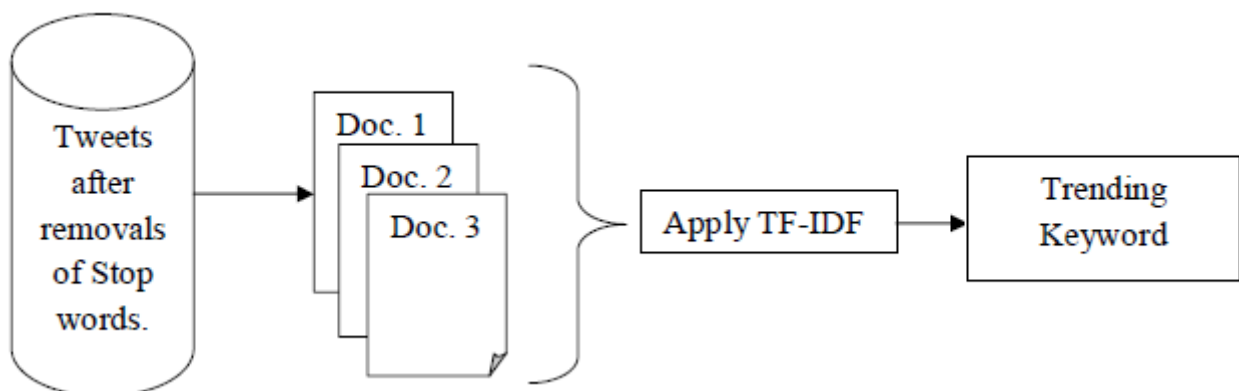

Figure 3 Grouping of Tweets and applying TF-IDF

We can find the trend by applying term frequency inverse document frequency on the data corpus. Here we have make different document of the tweets on the daily basis. For the time period of a month we will have 30 documents containing different users tweets. On these documents we will apply the text mining technique for trending keyword.

In our analysis the trending keywords for the time period of one month. In analysis we can see that Justin Bieber is the high on all time so it can be considered as Steady Trending Keyword. As there is also Daniel Sturridge which is spontaneous trending Keyword for one day. In mean time Daniel Sturridge joined Liverpool from Chelsea. And after that this keyword shows low rank on trending keywords and eliminated from the trending keywords list. From our proposed method we can find the trending topics and keyword from tweeter stream data.

V. CONCLUSION AND FUTURE WORK

Social network allows millions of users to shares and consumes its content. The Data generated from Social Network can be extracted and analyzed for Trend Identification. It is important for organization to know these trends as quickly as possible. Use of mobile for sharing information from mobile is increasing day by day. This leads the information shared on social network contains typing mistakes, abbreviations and synonyms. Solving these for the proper trend identification is necessary. In this thesis I introduced a system for trend identification on social network with ability to solve the typing mistakes, abbreviations and synonyms used for the trending information.

We can apply different text mining technique for trend identification like Cosine Similarity, Latent semantic indexing, Latent Dirichlet Allocation, nonnegative matrix factorization etc. The result set can be compared to this techniques and can be find out which method gives best result. Social network consist of media information also to be shared, the proposed system will work with text data, for more accurate results the media information can also be analyzed and approach can also be applied for different social networks.

REFERENCES

Papers

[1] D. Ramage, S. Dumais and D. Liebling, "Characterizing microblogs with topic models", in In Proceedings of the 4th international AAAI Conference on Weblogs and Social Media, (2010), pp. 130–137.

[2] G. Kumaran and J. Allan, "Text classification and named entities for new event detection", in Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, (2004), pp. 297–304

[3] M. Mathioudakis and N. Koudas, "TwitterMonitor: trend detection over the twitter stream", in Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, New York, NY, USA (2010), pp. 1155–1158.

[4] H. Sayyadi, M. Hurst and A. Maykov, "Event detection and tracking in social streams", in In Proceedings of the International Conference on Weblogs and Social Media, (2009).

[5] Daehoon Kim Daeyong Kim, "Detecting Trend and Bursty Keywords Using Characteristics of Twitter Stream Data ", in International Journal of Smart Home Vol. 7, No. 1, (2013), pp. 209–220.

Websites

[6] Wikipedia Guide http://en.wikipedia.org/wiki/Web_mining

[7] Google Trends, http://trends.google.com

[8] Tweeter Developer API, https://api.twitter.com/

[9] TweetInvi API for C#.NET, https://tweetinvi.codeplex.com/