Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 1): pp. 5155-5159
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.01.550

Real Time Event Detection From The Twitter Data Stream

Ms.V.Narmadha[1] , Ms.P.Padma[2] , A.Vignesh[3] , S.Naveen Krishna[4]

[1,2] Students, Department of Information Technology, Sri Sairam Engineering College

[3,4] Associate Professor , Sri Sairam Engineering College

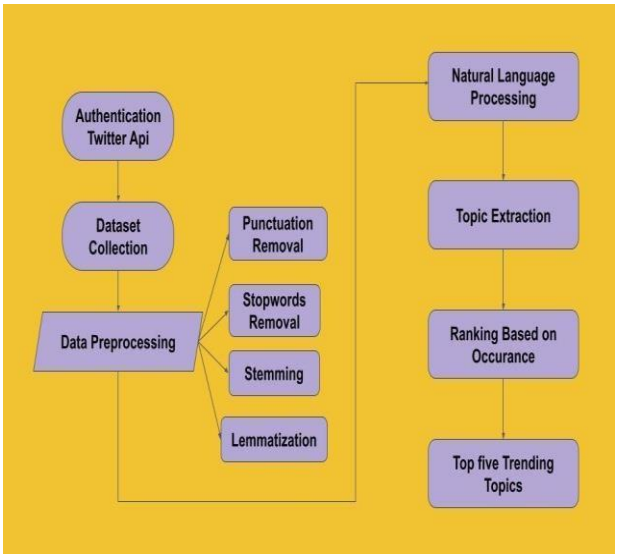
ABSTRACT:

The usage of popular social media like Twitter , facebook has increased in recent times. The trends in Twitter have the ability to promote many events such as political events, market changes and other types of breaking news.In our project, we propose to find those trending events without using the hashtag count. In the proposed system, the live tweets are streamed by using the Twitter Api.Those Tweets are preprocessed and the trending topics are extracted using Natural Language Processing. This process will show drastic changes in the trending topics as many people don't know the exact hash tags.

INTRODUCTION:

Social Media is a very power tool medium of information transfer. Every one ranging from children to elders use social media.Due to the exponential increase in the use of the internet by people of different cultures. The things which are getting trending in twitter is calculated by using the hash tags mentioned in the tweets. Many users don't know the exact hashtag which is to be used for their own tweet in the means of the event. Our project aims on collecting and ranking the top topics which are in trend by using the data mining process. We use the Natural Language Processing method to identify the particular topic in which the tweet has been posted and then collect them in a separate space so that we can rank them based on the repetition of the topic. Our aim is to reduce the error of popularity of events which may occur when using the hashtag methods.

ARCHITECTURE DIAGRAM:



Proposed System:

Dataset collection:

A collection of proper tweets based onlocation is taken as a dataset for the processing. A data set is a collection of data.The data set lists values for each ofthe variables, such as height and weight ofan object, for each member of the data set.Each value in the table is called a datum.Commonly the dataset consists of number values for the prediction process butin our project the tweets of the particular user in the format of sentences is acting as the data present in each row. Some commonattributes like the tweet ID , username and tweet time are also given as attributes in the dataset. The dataset here is kept in the format of comma separated values (csv) for easy access of data.

Twitter Api:

Twitter Api provides the feature of extracting the live tweets by using some authentication Id such as consumer key,consumer secret, access key and access secret.Using this the streamed tweets and thedata related to those tweets are stored in a csv file for further processing. Some of thekey parameters to get the tweets are location,time and language of the tweet.

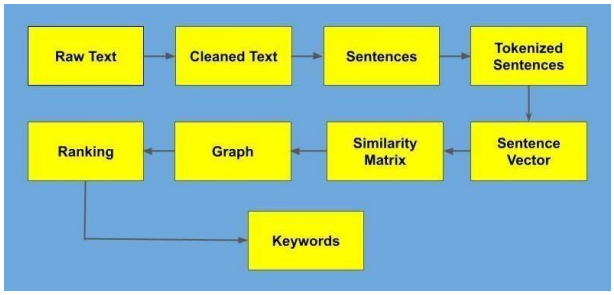
Data pre-processing:

Data preprocessing is a cleaning technique which is used to convert / transform the raw data into a clean and properly structured dataset suitable for further analysis. Data is usually collected and gathered from various sources , so it should be good enough and in some specific format before the model learns or gets trained with the data. This will help in achieving better

and accurate results with valuable information. The basicsteps in pre-processing involve filling up missing values and null values , getting rid of possible outliers and normalisation.The data is preprocessed in a way that each tweet is preprocessed into a list of keywords by using some basic preprocessing like Punctuation removal, Stopwords removal, Stemming and Lemmatization.

Processing of Data:

The processing of data is done by using NLP methods. Natural language processing (NLP) is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data. Here the keywordsare cleaned in an efficient manner then a sentence vector for each tweet is created. Asimilarity matrix is created so that similar words are arranged. A graph is plotted between the similar words in this method for ranking of data.Natural Language Processing:



ADVANTAGES OF NATURALLANGUAGE PROCESSING:

1. It processes the data in a relatively faster manner than theother methods.
2. The conference resolution ismuch higher in NLP.
3. The Discourse Analysis is donein a very efficient way.
4. It produces automatic summarization for the givenlarger data
5. It doesn't offer any unwanted orunnecessary information to the user.

Sentence Vectorization:

In regular sentences many noisy unwanted words may be found. As these types of dataare not meaningful and do not provide any information so it is mandatory to remove these types of noisy data.

Similarity Matrix:

A similarity matrix is a matrix of scores which express the similarity between two data points. The words which are extracted are plotted in a similarity matrix for the elimination of related words.

Similarity Graph:

Graphs are a powerful tool for many practical problems such as pattern recognition, shape analysis, image processing and data mining. A fundamental task in this context is that ofgraph matching. When computing the similarity (or its extensions) of large graphs, the complexity can still be quite high, since one needs to solve an eigenvalue problem of a dimension that is essentially the product of the number of nodes in both graphs.

The keywords thus obtained from the processing are ranked on the basis of occurrence and the top five topics which are obtained from the process are displayedas the result.

CONCLUSION:

The goal of our project is to find the top trending topics on twitter without the usageof Hashtags. Exploring the importance and complexity of utilizing social media in disaster relief operations, specifically usingthe micro-blogging site - Twitter. This paper also extends the future,if we add features like extracting the data based on particular topics like disaster management and social movement. The android
application for this project can also becreated.

REFERENCE:

[1]E.S. Tellez, S. Miranda-Jiménez, M. Graff, D. Moctezuma, O.S. Siordia, E.A.Villaseñor, A case study of spanish text transformations for twitter sentiment
analysis, Expert Syst. Appl. 81 (2017)
457–471.

[2]A comprehensive analysis of trending twitter topics , Issa Annamoradnejed , Jafar Habibi

[3]M. Martínez-Rojas, M. del Carmen Pardo-Ferreira, J.C. Rubio-Romero, Twitter asa tool for the management and analysis of emergency situations: A systematic literature review, Int. J. Inf. Manage. 43 (2018) 196–208.

[4]Pritam Guha Sentimental analysis on twitterdata regarding 2020 US election Nov 2020 .

[5]N. Öztürk, S. Ayvaz, Sentiment analysis on twitter: A text mining approach to the syrian refugee crisis, Telemat. Inform. 35 (1)(2018) 136–147.

[6]C. Diamantini, A. Mircoli, D. Potena, E. Storti, Social information discovery enhanced by sentiment analysis techniques, Future Gener. Comput. Syst.95 (2019) 816–828.

[7]'What are you Tweeting about?': A survey of Trending Topics within Twitter. Marc Cheong.