Role of Natural Language Processing in Social Media

Aditya Andhare[1*], Surabhi Thorat[2] and Vikrant Shaga[3]

[1*]Department of MCA, MIT(E), Aurangabad, India
[2]Department of MCA, MIT(E), Aurangabad, India
[3]Department of MCA, MIT(E), Aurangabad, India

[*]*Corresponding Author:* shagavik@gmail.com

Abstract— This paper highlight role of Natural Language Processing in social media sources like twitter, Facebook, LinkedIn etc. As now a day's social media playing the vital role in terms of current trends social issues awareness etc. The major area where social media analysis requires is business analytics. As every business need the customer reviews and preferences for their business growth. This paper explains the major steps involved in social media mining.

Keywords—Natural Language Processing, Social media,text summarization, sentiment analysis, Named Entity Recognition, Part-Of-Speech,Tagging.

I. INTRODUCTION

Social media plays a very important role in day to day life. This is the new trends in current technologies. We need to extract important part from free form of social media communication. NLP can analyze language patterns to understand text. One of the most addictive ways NLP offers valuable intelligence is by tracking sentiment.
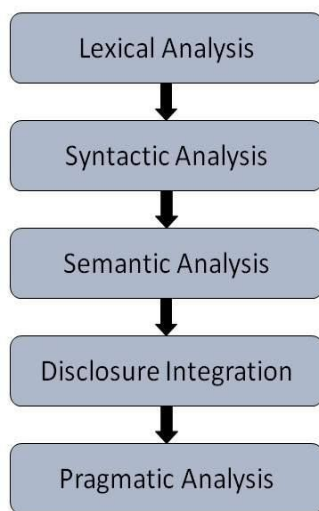

Fig.01 Stages of NLP [7]

This paper focuses majorly on a role of natural language processing in social media and its various approaches for analysing, processing text. Also this paper highlights the process of data collection applied during the processing of natural language.

II. NATURAL LANGUAGE PROCESSINNG

A. Text summerization

Text Summarization is a process of collecting information from original text to generate some result. In recent years the need of text summarization is can be seen in various fields such as summary of an email, news summery, field of science (Medical) etc. [1]. Another meaning for Text Summarization may be like reducing the large document without changing the real meaning of the document [2], that means it creates an summary of the text file. The summary which have generated is help user to fine related information regarding their search easily, it also gives overview of the written text, but it is getting very difficult to the user to find out the related information from this large and vast world of INTERNET so hence there is need of Text Summarization [3].

In this we look for the term which is occurring for the multiple times, which we called as Salient. We calculate the score of the sentence means we look for frequently occurred words [2].We find out the location of the text means we look for the text whether it is coming at the start or at the last sentence. Using this method we can find out the positivity and the negativity of the sentence or it tells the importance of the text. Such a cues are: "in summary", "in conclusion", "and hence" etc. It looks for the similarity between the sentences. Title, Headlines etc is the things which can be compared to check the similarity between the two sentences. It counts the total size of the summary. Most of the times very large and very small sentences are also not perfect for the summary [2].

B.   Named Entity Recognition

Named Entity Recognition helps us to identify entities or present objects such as names of an organization, name of person, places in the text. For example Named Entity Recognition of sports news may consist of names of player, name of the team, ground where they are playing etc.[4].

In many of the organization like Stanford University they use Named Entity Recognition as a solution to their NLP related problem. Stanford NER is stands for Named Entity Recognizer which is implemented on the JAVA platform [5].

C.   Sentiment analysis

The work of sentiment analysis is to identify, extract the attitude of the person. Text collection shows different sentiments which are positive, negative, or neutral. Sentiment analysis is mostly used in the surveys like online reviews or social media monitoring It identifies the sentiment and returns numeric value between 1.0 to -1.0 where 1.0 is fully positive sentiment and -1.0 is fully negative sentiment.[6].

In general Opinion's have two components: A target and sentiment towards target. Sentiment analysis mostly done after the text parsed by POS (Part of speech tagging). Simple example of sentiment rule is if two adjoining positive words adjective and adverb respectively then it is classified as Positive Sentiment [5].

D.   POS Tagging

In this NLP approach, sentences will be tagged according to grammatical forms which are nouns, verbs, adjectives, adverbs etc., but the problem is same words can have multiple forms such as noun and verb both. Dictionary contain multiple categories of words with same meaning and different meaning. Usually tagging words of sentences accurately tag the word or by making guess. In such cases POS tagger use probability approach on ambiguity creating words [4].

Social media mining

III.   DATA COLLECTION

In this step we collect only necessary data. We consider this data as a unimportant data because no valuable information is retrieved at this point. Mostly we use important keywords to find required data. All we need to do is gather historical data[4].

There are 'n' numbers of peoples who access historical data at specific time and it becomes very hard and costly for social networking to collect large amount of data. So the work of summarization keeps all important data. [6].

Different sources were we can collect the data for mining are [8]:

1) Blogs- Blog posts contains the thoughts   an individual.These blogs may contain reviews on many products, issues related to social economical and current trends etc.

2) Review Sites- These sites are highly in demand for business analytics and there growth. The data given by reviewers are collected from the e-commerce websites like flipkart,myntra etc.

3) Research dataset- The dataset is available by many research university dataset portal for research work that contains different types of product reviews extracted from Amazon.com including movie view, mobile review etc.

4) Microblogging sites- It is one of the most commonly used communication tool among Internet users is micro-blogging. Millions of messages appear daily in popular web-sites for micro-blogging such as Twitter, Facebook.

*1)*Pre-processing-Processing of raw data will provide platform for data analysis. The main reason of pre-processing is to converting raw sentences into machine readable format. User enters social text in free form so hence it is difficult task to differentiate that data. To do sorting out from this challenge we are having techniques like Part-of-speech and Named entity recognition are used.[6] This approach is very costlier than other technologies in terms of storage needed for storing documents. For accurate interpretation on the data this approach is very useful. [4]

*2)*Mining process-This is the main part in which we extract actual important information from the classified data. Frequent pattern mining is the task which we perform in this step.

We use this step to identify frequent behaviour of an object or entities. Firstly this method was used for identifying customers buying behaviour. Another application consists of opinion mining is to find opinions about the product. If we take the example of movie review it will illustrates it clearly.[4]

IV.   CONCLUSION

This paper explains the impact of social media in business, social and economic issues, how social media can be analysed through natural language processing to observe the sentiments of people. In this paper we tried to high lighten major concepts related to natural language processing .

V.   REFERENCES

[1]   https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_natural_language_processing.htm

[2]   Saranyamol C S, Sindhu L, "A Survey on Automatic Text Summarization", International Journal of Computer Science and Information Technologies, 2014,Vol. 5 Issue 6.

[3]   Rafael Ferreira et al."Assessing Sentence Scoring Techniques for Extractive Text Summarization", Elsevier Ltd., Expert Systems with Applications 40 (2013)5755-5764

[4]   H. P. Luhn, "The automatic creation of literature abstracts", IBM Journal of Research and Development, vol. 2, pp. 159-165, 1958.

[5]   John Selvadurai "A Natural Language Processing based Web Mining System for Social Media Analysis", International Journal of Scientific and Research Publications, Volume 3, Issue 1, January 2013 ISSN 2250-3153.

[6]   The Stanford Natural Language Processing Group. (2012, Nov 15). Stanford Named Entity recognizer (NER) [Online].

[7]   P. Radha and R.Divya," Extracting Information from Social Network using NLP", International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 4 (2017), pp. 621-630

[8]   Surabhi Thorat," Opinion Mining and Sentiment analysis- Its Tools and Challenges", International Journal of IT, engineering and Applied Sciences Research (IJIEASR) ISSN: 2319-4413 Volume 3, No. 11, November 2014, pp. 12-15