NLP in social Networking: An Overview

Amit Kumar
(Department of CSA, F.L.T.M.S.B.P GCW, Rewari, Haryana)

ABSTRACT: *Today social network is so popular. So Analysis in this area is so important .social Network analysis refers to the process of discover distributed structures through the use of network and graphs. All the information on the social network is unstructured and we have need to get the structured information that will helpful for making use of the important information. Natural language processing improve the accuracy in envisage the structured information that is speckled over social network. The basic overview of NLP and its technique are showing in this paper. This paper shows various text mining systems which is the foundation of NLP to analyze social network information.*

INTRODUCATION

In this modern era **,**everything is conducted online from the creation of the product to its marketing. It is also beneficial for business organizations to build relationship on mutual sharing, start and stop sailing ,get review of customers they are satisfied from the product or not. Social Network is useable almost in every field like voting mechanism for beauty pageant, political campaigns, medical research, scientific research and promotion via advertisements. We further need for analyzing and modeling of such networks. We can inquire more and also can compare about organization's external environment with the help of competitive intelligence. Data mining techniques are needed for handling the three properties with social network which are size, noise and dynamism. This property requires automation for dispensation of information, analyzing it within a fixed time. Interestingly, data mining techniques are designed to handle the large amount of data. So data mining is very useful for social network to make it user friendly.

NATURAL LANGUAGE PROCESSING (NLP):

This paper analyses excessive use of Natural Language Processing and web mining techniques to study Social Network. **Neuro-linguistic programming** (**NLP**) is an approach to communication between human and machine language created by Richard Bandler and John Grinder in California, United States in the 1970s. NLP is basically used for information retrieval, word prediction ,spell and grammar checking, text categorization, This paper discusses NLP approaches that are essential for Social Networking these are Summarization, Chunking, Part-of-speech tagging, Named Entity Recognition, Named Entity Disambiguation, Fact/Relation Extraction etc.

A) Summarization:

Summarization is the process of creating a condensed form of text document which maintains significant information and general meaning of source text or we can say that generating a short summary from one or more documents, sometimes based on a given query. There are two types of approaches for automatic summarization [6]: Extraction and Abstraction. Extraction refers to selecting a subset of given words or sentences in the original text to form the summary and abstraction that builds an internal semantic representation and then use natural language generation techniques to create a summary that is closer to what a Human might generate.

B) Parts-of-Speech Tagging:

The process of assigning one of the parts of speech to the given word is called parts of speech tagging. It's commonly referred to as POS tagging. Parts of speech include nouns, verb, adverb, pronouns, conjunction and sub categories. For example word: Go tag: verb here it will show go is a verb.

C) Chunking:

Chunking is the basic technique used for entity detection. Chunking is a cluster of information or group of information in the direction of a Deductive or Inductive conclusion through the use of language. There are two type of chunking 1. Up chunking 2. Down chunking

D) Named Entity Recognition:

Named Entity *Recognition* is a subset of information extraction that locates and classifies Named Entities in text into pre-defined categories such as the names of persons, company, locations, quantities, monetary values, percentages, etc.

E) Word Sense Disambiguation

The word sense disambiguation technique is use to identifies the correct sense of the word in a sentence where multiple meanings of the word exist. This methodology provides a mechanism to diminish the ambiguities of words in the text [1][2]. For example 1.I went fishing for some sea bass.2.The bass line of the song is too weak.
In this example bass has two meaning i.e bass is a type of fish & other one is the tone of frequency so

this technique sense the meaning of bass in both sentence with different way.

F) Fact/Relation Extraction:
Once named entities have been identified in a text, we can then extract the Text and relations facts that exist between specified types of named entity. The objective of the fact extraction is to detect and differincate the semantic relations between entities in text and relations and fill it in a predefined template using the entities

APPICATION OF NLP:
Some basic application of NLP are: spelling and grammar checking ,optical character recognition, screen reader for blind and partially sighted user, information retrieval, document clustering, question answering, text segmentation, exam marking, machine translation etc.

CHALLENGES IN NLP
1.Informal language: Social Network users Post texts in an informal language which is noisy include lack of misspellings and grammatically corrections.
2. **Part-Of-Speech** : Tags make the Information Extraction from social network more challenging.
3. **Noisy sparse contents**: The users' post on social network does not always contain useful information. To purify the input posts stream, Filtering is required as a initial step.
4. **Short contexts**: It is difficult to disambiguate mentioned entities due to the briefness of the posts and to resolve co-references among the feeds.
5. **Information about entities**: People normally use social Network to express Information about their daily routine so the entities are not contained in the knowledge Base.

TEXT MINING
Text mining refers to the automatically extracting information from a usually large amount of unstructured resources.NLP used text mining to extract Meaningful information from free text Searching with the help of text mining is a way of retrieving and searching information through social network like images, news, videos etc.
There are four steps in the text mining, which are:
Data Collection, Preprocessing, Generalization and Analysis
A) Data Collection:
Data collection is the process of information gathering and measuring in a systematic
way, which then give answer to the relevant questions and evaluate Outcomes. There are large numbers of users who access historical data at a particular time and it becomes difficult and costly for social network to gather large amount of data. So,

Summarization maintains all important data and further discards the irrelevant data.
B) Pre-processing:
Preprocessing is an important task and critical step in Text mining, Natural Language Processing (NLP) and information retrieval (IR). In the area of Text Mining, data preprocessing used for extracting interesting and non-trivial and knowledge from unstructured text data. Information Retrieval (IR) is essentially a matter of deciding which documents in a collection should be retrieved to satisfy a user's need for information. Users enter the social text in a free form and therefore it is a challenging task to classify that data. Just to be sorted out from this challenge, part-of -speech tagging and Named Entity Recognition are used [3]. Pre-processing consist of following step i.e Tokenization, Filtering, Lemmatization,   Stemming.
C) Generalization:
Generalization involves the multiple patterns at the text of the preprocessed texts. It
Deals with developing algorithms to ascertain stimulating, unforeseen and
Unusual information from the patterns in the text document. One of the common tasks that occur is referred to as Apriori [4].
D) Analysis
It deals with the validation and .understanding of the generalized data .Here we deals with
Density, Centrality, indegree, outdegree terminologies to analyze the social network[6] .Here Degree identifies the "connections" between the users.

Applications of Text Mining in Social Network
Some applications of text mining in social network are[5]:
1.**Keyword Search**
In keyword search, we specify a set of keywords, which are used to determine social network nodes which are relevant to a given query. In the problem of keyword search, we use both the content and the link behavior in order to perform the search. Query semantics, query efficiency are major concern of keyword search.
2. Classification
 In classification, in the social network the nodes are associated with labels. These labeled nodes are then used for classifying the social network. There are various algorithms are available for classification of text from content.
3. **Clustering**
In clustering, we determine set of node having similar content for clustering. There are various technique and algorithm for clustering. Clustering is widely

used in classification , visualization and document organization .

5. CONCLUSION

This paper discussed Basic overview about NLP and NLP techniques for Social Network that can improve the experience of the user in meaningful way.

REFERENCES

[1] Bandyopadhyay, S. (Ed.). (2012). Emerging Applications of Natural Language Processing: Concepts and New Research: Concepts and New Research. IGI Global.

[2] Princeton University. (2012, Nov 10). WordNet: A Lexical database for English [Online]. Available: http://word net.princeton.edu/

[3] Bandyopadhyay, S. (Ed.). (2012). Emerging Applications of Natural Language Processing: Concepts and New Research: Concepts and New Research. IGI Global.

[4] Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings of the 20th VLDB conference, pp 487–499

[5] Aggarwal, C. C., & Zhai, C. (Eds.). (2012). Mining text data. Springer Science & Business Media.

[6] Bikel, D., & Zitouni, I. (2012). *Multilingual natural language processing applications: from theory to practice*. IBM Press pp 400.

[7] Xiang, R., Neville, J., & Rogati, M. (2010, April). Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web* (pp. 981-990). ACM.

[8] Louis, A. (2017). Natural Language Processing for Social Media

[9] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

[10] Kasemsap, K. (2016). Text Mining: Current Trends and Applications. Web Data Mining and the Development of Knowledge-Based Decision Support Systems, 338.

[11] Lin, J., & Dyer, C. (2010). Data-intensive text processing with Map Reduce. Synthesis Lectures on Human Language Technologies, 3(1), 1-177.

[12] M. Song and Y. Wu, Handbook of Research on Text and Web Mining Technology, Hershey, PA, USA: IGI Global, 2009, pp. 228

[13] Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, *1*(1), 60-76.