

NATURAL LANGUAGE PROCESSING AND WEB MINING IN SOCIAL MEDIA ANALYSIS

VC.Agilarani^{1#},M.Tech(CSE),P.Karthika^{2#},M.Tech(CSE),J.Pamina^{3#}

#Student, School of Computing, KARE,vc.agilarani@gmail.com

#Student, School of Computing, KARE,karthikaagila@gmail.com

#Professor, School of Computing, KARE,jpamina8@gmail.com

ABSTRACT

Natural language processing is a way to provide human and computer communication and it is a processing of text or speech. Social media is an emerging technology that contains billions of users all over the world. Analyzing the data in social media is area of research and development that includes big data analytics, artificial intelligence, network analysis and natural processing. This paper deals with the study of natural language processing to analyze the data in social media.

KEYWORDS: Natural Language Processing, Web Mining, Social Media Analysis.

1.INTRODUCTION

Social media allows customer to provide opinions or reviews about the products and services. Social media provides a customer and consumer interaction. Social media monitoring systems is the emerging research technology. This paper discusses about natural language processing and web mining techniques for social media analysis.

2.NATURAL LANGUAGE PROCESSING

Natural language processing is a way to provide an efficient way of human and computer interaction. The main aim of natural language Natural language processing bridges the gap between human and computer using Information Extraction, Summarization, Question and answering and Clustering.

2.1 Approaches used in NLP

The approaches used in NLP are: Automatic Summarization, Named Entity Recognition,

Part of speech tagging, Word-Sense Disambiguation, Sentiment Analysis

2.2 Automatic Summarization

Automatic Summarization is a way to reduce large text into short paragraph. There are two approaches used in Automatic Summarization are Extraction that extracts keyword from sentences or paragraphs and Abstraction which paraphrases the important key points of the text. Analysis, Transformation and realization are the steps involved in automatic summarization.

2.3 Named Entity Recognition

Names of persons, Organization and Places in the text are identifies as entities or objects is called Named Entity recognition.

2.4 Part of speech tagging

Lexicons, Rules and Dictionaries are the techniques used by Part-of-Speech taggers. Nouns, verbs, adjectives are the words used to tag the sentence in a grammatical order

2.5 Word-sense Disambiguation

This technique allows to find the exact meaning where the sentence or word have multiple meanings. It helps to minimize the ambiguities of words in the text.

2.6 Sentiment Analysis

The way to identify the attitude or contextual polarity of the writer is sentiment analysis. Online reviews for products, movies and books can be done using sentiment analysis. Sentiments can be positive, negative or neutral

3.WEB MINING

The process of extracting or retrieving the information and patterns from web is called Web mining. Web mining is a combination of web structure mining, web usage mining, web content mining. Web structure mining deals with link structure of several web sites. Web usage mining includes three steps they are preprocessing, knowledge discovery and pattern analysis. Web content mining provides useful knowledge from webpage contents.

Classification, Association, clustering and personalization are the techniques involved in web mining techniques. Data gathering, preprocessing, indexing and mining are the steps involved in web mining process

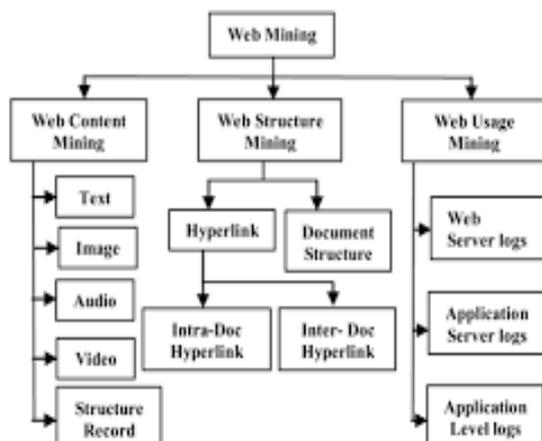


Fig1: Classification of Web Mining

4. SOCIAL MEDIA ANALYSIS

Collection of textual information in the web is called social media. Social media analysis is a way of analyzing the social networks (Facebook, Googleplus, LinkedIn), Blogs and

comments (Online Journal, Business Insider), microblogs(Twitter),Forums(OnlineDiscussion Communities),SocialBookmarks(Pinterest,Googlebookmark),Wikis(wikipedia) Social news (Reddit) and Media sharing (youtube, instagram).The main focus is on NLP applications, tools and methods that integrate linguistic information in social media analysis.

4.1 Data Collection from Social Media

Social data collection varies based on tasks and applications. Using social media service application programming interface other applications will be integrated.

4.2 Spam and Noise Detection

Spam detection provides online opinions such as fake positive and fake negative reviews and it also avoid damage to reputations. Social Media noise is analyzed by applying linguistic, statistical analysis and lexical composition.

4.3 Privacy in social media analysis

An interactive relationship between individuals, organizations and societies plays an important role n social media and privacy is a major concern in a social media. Privacy in social media involves user misunderstanding, bug detection, unauthorized access, lack of ethics in marketing.

5. RELATED WORKS

In [2] Samta Tembhekar presented the approaches used in natural language processing they are distributional ,frame based, model theoretical and interactive learning .It also deals with steps involved in NLP and terminologies used in NLP. In [4] Raymond describes about web mining categories such as web structure mining, web content mining, web usage mining and information retrieval. In[7] XinChen discuss about the qualitative analysis and large scale data mining according to the engineering students twitter posts. It also deals with positive and negative reviews.

6. CONCLUSION

This paper discusses about the survey of natural language processing techniques and

web mining techniques to monitor the data in social media analysis. Combination of both natural language processing and web mining steps gives better result. It is used to analyze the positive and negative reviews of the products and services and comments on posts in social media

REFERENCE

[1]John Selvadurai (2013) A Natural language processing based web mining for social media analysis, International Journal of Scientific and Research Publications, Volume 3, Issue 1.

[2]Samta Tembhekar (2017) A survey paper on approaches of Natural Language processing ,WWW.ijariit.com,volume3,Issue 3

[3]Bharate and D. Gadekar (2014) Survey Paper On Natural Language Processing www.ijcea.com ISSN 2321-3469 Volume VIII, Issue III.

[4].Raymond Kosala ,Web mining Research- A survey, SIGKDD Explorations ,volume 1, Issue 1

[5]Geeta R. Bharamagoudar (2012) Literature Survey on Web Mining, IOSR Journal of ComputerEngineering (IOSRJCE)ISSN: 2278-0661, ISBN: 2278-8727 Volume 5, Issue 4

[6]Rohit John Joseph A Semantic Approach to Classifying Twitter Users ,Springer

[7]Xinchen (2014)Mining Social Media Data for Understanding Students' Learning Experiences, IEEE Transactions On Learning Technologies, VOL. 7, NO. 3

[8]Atefeh Farzindar, Natural Language for social media.