IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10598

A Survey on Intelligent Data Mining of Social Media for Better Decision Making in Product or Services

Niraj Shinde, Tejas Kulkarni, Ojas Nampurkar, Shraddha Payake, Dr.S.N.Gujar

Tssm's Bhivrabai Sawant College of Engineering and Research Narhe, Pune

Abstract: The emergence of social media and the quick growth of mobile technologies of communication have radically modified the manner of expressing emotions, mood, enthusiasm etc. People typically use brief epigrammatic writings to communicate their reactions, envies and preferences via social media rather than to create large letters. Many websites like Twitter, Google Review, Just Dial, book my Show, etc. allow individuals to share and discuss thoughts, opinions and views in the form of a short text that can be used to make decisions about whether or not such service or product is good to other unknown people, customers and service users. This paper divides the entire process into two phases. In the first step intelligent data mining data will be abstracted; and in the second stage frameworks will be developed to focus on positive and negative opinions and to assist people decide on their issue using R-programming (i.e., on institute, services, products, movies, tourist spot etc.). It contains graphs, images, and so forth in the visualization area. It can also incorporate a comparison parameter which will again be of great use to users and peoples.

Keywords: Co-extracting algorithm, co-extracting model, opinion targets, Opinion Relation Graph, opinion words, Topical Word Trigger Model.

I. INTRODUCTION

Online networking provides many opportunities for patients to communicate about their encounters with medicines and devices, as well as for organisations. In their IT offices, pharmaceutical organisations organise informal inspections, create open doors for the speedy dissemination, supply and improve transmission services for articles and administrations, increase turnover and benefit, and save expenses. In addition, the collection of online information for bio-monitoring was accounted for. The idea of informal communities makes information gathering troublesome. A few techniques have been utilized, for example, link mining, classification through links, predictions based on objects, links, presence, estimation, protest, aggregate, and subgroup location, and mining the information. Connection forecast, viral showcasing, online talk gatherings (and rankings) take into consideration the advancement of arrangements in view of client criticism.

We utilize the Self Organizing maps to detect the correlations of users' postings with a good or unfavorable view on the medicine during the first stage of our present research. We shape the users and their postings utilizing a network-based strategy in the second phase. Manual qualitative examination of a wide range of data related to software in order to identify the data values of the postings of software users. To properly record and classify the many forms of actionable software maintenance requests in such posts, use Text classification algorithms. To study the performance of various text resuming approaches in the production of short summaries of the common technique. Our fundamental goal is to create a more responsive and flexible software engineering infrastructure that can effectively and quickly deliver customer delight.

LITERATURE SURVEY

1. **Artificial Societies and Social Simulation Using Ant Colony, Particle Swarm Optimization and Cultural Algorithms,** This system proposes Ant Colony System Algorithm. Artificial Societies and Social Simulation using different strategies to analyze and model the necessary information to support the correct decisions of the evolving models. Advantages: Improves good quality in a short time. It has better performance. Disadvantage: Community of agents is not in application. [1]
2. **Collective Extraction for Opinion Targets and Opinion Words from Online Reviews,** proposes a method to extract opinion targets and opinion words collectively based on the word alignment model. a collective extraction for opinion targets and opinion words based on the word alignment model, in which the extraction can be treated as a classification problem. Design a semi-supervised extraction method based on active learning, since labeling training samples is time-consuming and error-prone. Advantages: Higher accuracy. Effectively ignore the problem of error propagation. Greatly reduce the work of manually labeling samples. Disadvantage: It does not apply parallel extraction method. [2]

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10598

3. **Tracing Information Flow and Analyzing the Effects of Incomplete Data in social media,** proposes a k-tree model of cascades is generated from a balanced tree of height and branching factor. The goal of this paper is to address methods to collect massive amounts of social media data and what techniques can be used for correcting for the effects and biases arising from incomplete and missing data. Advantages: The information flow is unambiguous and precise. We can have the time, so it's easy to trace the information. In a very large network, it becomes easy to collect data. However, if data is incomplete cascades break into pieces. Many different diffusion mechanisms. Disadvantage: Not all the links transmits the information. Sometimes the links get missing due to Blogger forget to attach a link or mainstream media does not provide the source links. Not clear whether hashtags really diffuse. Due to "personalization" easier to argue URLs diffuse. Problem with all is that we do not know the "influencer". [3]

4. **Text and Structural Data Mining of Influenza Mentions in Web and Social Media,** Proposes a graph-based data mining technique to detect anomalies and informative substructures among flu blogs connected by publisher type, links, and user-tags. Text mining of influenza mentions in WSM is shown to identify trends in flu posts that correlate to real-world ILI patient reporting data. Advantages: To identify trends in flu posts that correlate to real-world ILI patient reporting data. Disadvantage: Content analysis does not provide. [4]

5. **Text Mining: Promises And Challenges,** Proposes a text mining framework consists of two frameworks: Text refining and Knowledge distillation. The text refining that transforms unstructured text documents into an intermediate form; and knowledge distillation that deduces patterns or knowledge from the intermediate form. Advantages: Customer profile analysis, Patent analysis, Information dissemination and Company resource planning. Disadvantage: There are issues in this paper, semantic analysis, multilingual text refining, domain knowledge integration and personalized autonomous mining. [5]

6. **Social media competitive analysis and text mining: A case study in the pizza industry,** Proposes competitive analysis for the user-generated data on Twitter and Facebook in three major pizza chains. Results from the text mining and social media competitive analysis show that these pizza chains actively engaged their customers in social media such as Twitter and Facebook. Advantages: Establishing effective and realistic benchmarks. Mining the content of social media conversations. Disadvantage: Does not track real-time data. [6]

7. **Mood Based Classification of Music by Analyzing Lyrical Data Using Text Mining,** Proposes classification using Support Vector Machine algorithm. As mood classes in music mental models may do not have the social connection of today's music listening environment, this research inferred an arrangement of mood classifications from social labels utilizing etymological assets and human skill. The resultant mood classes were contrasted with two delegate models in music brain science. Advantages: The framework may be utilized to hunt down female craftsmen, content melodies, or hallucinogenic music. Content features may prompt higher correctness's for most mood classifications. [7]

8. **Text Mining for the Hotel Industry,** This paper proposes text mining as a means of information management. Text mining can analyze the voluminous textual information that can be found in a hotel's internal databases and external sources. In this paper, illustrates how the text-mining technique may help to translate online textual information into meaningful competitive and customer intelligence for managerial decision making. Advantages: Reduce the use of manual labor in identification, storage, and analysis of business intelligence. Fully automated system. Disadvantage: Does not integrate text-mining tools with related technologies such as image recognition and Web-table mining. [8]

9. **Feature extraction and classification of proteomics data using stationary wavelet transform and naïve Bayes classifier,** Proposes Naïve Bayes Algorithm and stationary wavelet transformation. The data processes of MS signal in this paper mainly include two parts: preprocessing and biomarker selection, and the results are determined mainly by these two steps. To the denoising using SWT, compared to DWT, SWT it is very appropriate for this application for the characteristics of the MS data. Advantages: It requires a small amount of training data to estimate the parameters necessary for classification. High sensitivity, specificity and accuracy. [9]

10. **Semantic Data Analysis Algorithms Supporting Decision-making Processes,** Proposes semantic data analysis processes, and theirs role in supporting decision-making tasks as well as intelligent management. The most important is that such systems may support financial or economy processes taken in different enterprises or institution. Wide information records obtained thanks to the application of cognitive information systems allow finding many different applications both in local and global environment. Advantages: Cognitive systems are very efficient. Disadvantages: The structure of information record is too complex to perform full interpretation. The system has not enough knowledge to fully describe the semantic meaning of analyzed information record or complex structure.[10]

II. PROPOSED SYSTEM

The aim is to detect where opinion, review or tweets are set up in reports. Extraction of feelings In words, phrases and documents, opinion is concealed. The least complete semantical unit from which views are deleted is a sentiment sentence. The words of opinion, the opinions and the logical data should be taken for granted, as the phrases and the inclinations are separated. Then the extraction algorithm is produced by first acknowledgment of opinion words and then shortly afterwards distinguishes between the sensational polarities of the phrases. In order to obtain the required data

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 5, May 2021

DOI 10.17148/IJARCCE.2021.10598

from software tweets, we used the Twitter Search API. We limit our data collecting technique in our research to tweets directed directly to a certain software system's Twitter account. We used the Twitter *Search API,* to collect our dataset of software relevant tweets. We examine the performance of the Sentiment Analytical Algorithm to categorizes our data automatically. In our analysis, we limit our data collection process to tweets addressed directly to the Twitter account of a given software system. To automatically classify our data, we investigate the performance of sentiment analysis algorithm.

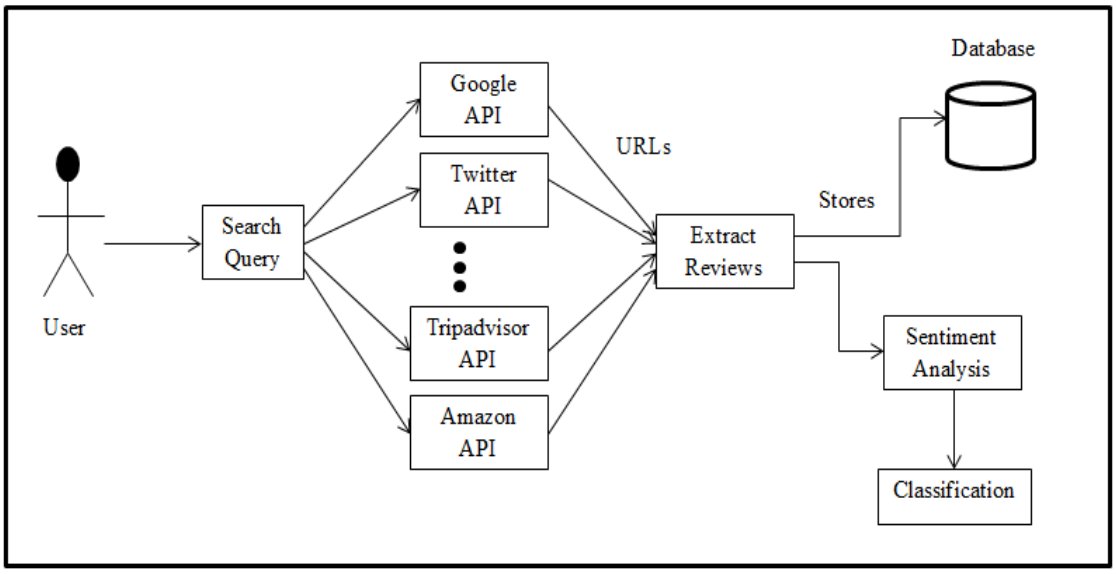

Fig.1 Proposed system architecture

Advantages:

- Less time consumption.
- Easy access.
- Pervasiveness.
- Data from different websites can be found at one place.
- Data scraping only single click button.
- Automatic classification.

III. CONCLUSION

This project is implement using data scraping technique. The number of posts is reached via data scraping. Apply the feeling analysis to each post, so that the user decides which one is excellent. This project decreases search time and labor on numerous websites. This offers users an easy access platform to make smarter choices. Added additional decision-making efficiency in the visualization and comparison. Globalization can be made available in future around the world. In the project, data may be extracted from different social media.

REFERENCES

[1]. Artificial Societies and Social Simulation Using Ant Colony, Particle Swarm Optimization and Cultural Algorithms
[2]. Collective Extraction for Opinion Targets and Opinion Words from Online Reviews
[3]. Tracing Information Flow and Analyzing the Effects of Incomplete Data in Social Media
[4]. Text and Structural Data Mining of Influenza Mentions in Web and Social Media
[5]. TEXT MINING: PROMISES AND CHALLENGES
[6]. Social media competitive analysis and text mining: A case study in the pizza industry
[7]. Mood Based Classification of Music by Analyzing Lyrical Data Using Text Mining
[8]. Text Mining for the Hotel Industry
[9]. Feature extraction and classification of proteomics data using stationary wavelet transform and naïve Bayes classifier
[10]. Semantic Data Analysis Algorithms Supporting Decision-making Processes