



Reviews on Opinions Mining and Summarization Methods for Analysis of Unstructured Textual Data for Decision Making

Shatakshi Agrawal, Dr. Sadhna Mishra, Prof. Gaurav

Computer Science & Engineering, LNCTS, RGPV,
Bhopal, India

Abstract— *the advancement and growing popularity in the computing facilities like Internet, social websites, E-commerce platform and many more have provided the people a new medium of sharing their views and opinions with each other in a very easy manner. The Opinions expressed by the peoples about a particular products or organization contains important information that is very helpful to the organization for their future action on improvement and development. This paper presents an approach towards the analysis of freely expressed people's opinions as well as different existing approaches of opinion mining are summarized in the literature reviews for finding the challenges and scope that are available in the field of opinion mining & analysis for efficient decision making.*

Keywords—*Opinion mining, Semantic Analysis, unstructured text data, words polarity.*

I. INTRODUCTION

Internet has gained a wide popularity now a day's due to the availability of information on button pressed sitting anywhere. The internet have provided users with free space for sharing information, suggestions, complain about the particular products, company, educational institutes, government policies and many more. The information or People's opinions are present in the form of freely expressed unstructured textual data like blogs, reviews, comments etc. The Opinions expressed in the blog post social networking websites present a wealth of information on the product and services if these opinions are carefully mined and summarized they can provide the vendors highly valuable information that helps in decision making and will facilitate him in growing the business. As a result Opinions Mining and summarization has recently received a great deal of importance within the researchers.

Opining Mining basically deals with finding the people's behaviour or action in response to a particular action. It means extracting the polarity about the entity from freely expressed textual data. Basically polarities are considered into many forms like positive polarity, negative polarity, very positive polarity, very negative polarity and neutral polarity. The polarity expressed about the particular aspects or features of the product or organization are summarized and based on that the results are formulated. The Summarization of polarity results helps in decision making for upgrading or for launching new policies.

The Opinions can be taken from the organization employees or from the students of the educational institutes for finding the problems and based on the opinion mining and summarization the problems can be easily identified. Because the students or employee have express their opinion freely in their own language about the organization in which they are working or studying. The collection of the data can be done by various traditional methods like pen and paper or by advance computing methods like websites and all but the problem arises that how to analyses this data. The analysis of these data in not an easy task various difficulties are present while analysing these freely expressed unstructured textual data.

This paper presents a review analysis on existing methods that can be utilized in order to solve the challenges that are presents while analysing the unstructured textual data. Section I deals with the introduction about the opinion mining. Section II presents an extensive literature survey about the methods and techniques for opinion mining. Section III focuses on the challenges that have come up by analysing & summarizing the literature survey. Section IV presents the proposed work for the system. The conclusion and future plan are stated in Section IV of this paper.

II. LITERATURE SURVEY

Katerina Kabassi & etal researches on multi criteria decision-making theories with a cognitive theory called human plausible reasoning (HPR) to provide personalized assistance via graphical user interfaces (GUIs). A GUI called intelligent file manipulator (IFM) helps with organizing computer file storage. The system reasons about user actions, goals, plans, and possible errors and offers automatic assistance in case of a problematic situation. Three multi criteria decision-making theories [simple additive weighting, multi attribute utility theory, and data envelopment analysis] were adapted, implemented, and combined with HPR, in turn. This process resulted in three different versions of IFM that were evaluated. [1]

Yingcai Wu & etal develop an opinion diffusion model to approximate opinion propagation among Twitter users. Accordingly, they design an opinion flow visualization that combines a Sankey graph with a tailored density map in one

view to visually convey diffusion of opinions among many users. A stacked tree is used to allow analysts to select topics of interest at different levels. The stacked tree is synchronized with the opinion flow visualization to help users examine and compare diffusion patterns across topics. [2]

Desheng Dash Wu et al develops methodology that integrates popular sentiment analysis into machine learning approaches based on support vector machine and generalized autoregressive conditional heteroskedasticity modelling. A corpus of financial review data was collected. Computational results show that the statistical machine learning approach has higher classification accuracy than that of the semantic approach. [3]

Fuji Ren, et al focus on a challenging problem of predicting users' opinions toward topics they had not directly given yet, which they define as user-topic opinion prediction. The main contributions of this paper are as follows: 1) Different from previous work recognizing emotional states/sentiments from online micro blogging data but ignoring whose they are, seek to find out who has what opinion of a specific topic in advance. Predicting individual's feeling about a given target is important for affective computing studies and able to be used to various applications. 2) To provide a solution, the author considers the opinion homophily among Twitter social friends and users' opinion consistency on content-related topics, and formulate them as social context and topical context mathematically. 3) Utilizing the learned emotional knowledge from the observed tweets and the social and topical context information, we propose a framework ScTcMF to predict the unknown user-topic opinions.[4]

Chien-Liang Liu, et al proposed and developed a movie rating and review summarization system by using the mobile environment. The author proposed system uses a latent semantic analysis (LSA) to identify product features in the movie reviews. The feature based summarization is important for identifying the features present. Statistical approaches were used for finding the polarity of the words. The author has uses an SVM machine learning approach for analysing and classifying the reviews. The LSA approach was used to reduce the size of the summary based on the user's preferred aspect.[5]

Xiaohui Yu et al presents explored the predictive power of reviews using the movie domain as a case study, and studied the problem of predicting sales performance using sentiment information mined from reviews. The author has approached this problem as a domain-driven task, and managed to synthesize human intelligence (e.g., identifying important characteristics of movie reviews), domain intelligence (e.g., the knowledge of the "seasonality" of box office revenues), and network intelligence (e.g., online reviews posted by moviegoers). The outcome of the proposed models leads to actionable knowledge that can readily be employed by decision makers. [6]

Jingbo Zhu, et al presents aspect-based opinion polling from unlabeled free-form textual customer reviews without requiring customers to answer any questions. First, a multi-aspect bootstrapping method is proposed to learn aspect-related terms of each aspect that are used for aspect identification. Second, an aspect-based segmentation model is proposed to segment a multi-aspect sentence into multiple single-aspect units as basic units for opinion polling. Finally, an aspect based opinion polling algorithm is presented in detail. Experiments on real Chinese restaurant reviews demonstrated that our approach can achieve 75.5 percent accuracy in aspect-based opinion polling tasks.[7]

Nathalie Camelin, et al describes a system for automatic opinion analysis from spoken messages collected in the context of a user satisfaction survey. A strategy was proposed for estimating opinion proportions from ASR results. Opinion-specific language models were used for ASR decoding to improve the detection of speech segments expressing opinions and for generating topic and polarity hypotheses in these segments. Methods are proposed for accepting or rejecting segments from messages that are not reliably analysed due to the limitations of automatic speech recognition processes, for assigning opinion hypotheses to segments and for evaluating hypothesis opinion proportions. [8]

III. CHALLENGES AND SCOPE

The different approach presents in the literature survey have provided with the information that the freely expressed textual data deals with some of the challenges.

- The opinions are in freely expressed text form so they are hard to structure and draw conclusions.
- Advanced user of statistical and data mining techniques are required in order to retrieve the information from the text.
- Advanced automation techniques are necessary for opinion mining and summarization of enormous textual data.
- The analysis of textual opinions is a major problem that needs to be addressed in order to retrieve valuable information for the organization.

IV. PROPOSED WORK

The basic idea of the proposed system to develop natural language based Opinions Mining and Summarization system for Unstructured Textual Data for extracting the valuable information for Decision as well as it will be capable of deciding the opinions polarity of those particular Features of the institute from the textual data.

The proposed system uses Natural language based approach for semantic datasets and polarity datasets generation. The student's opinions about their educational institutes will be considered in order to test the working and effectiveness of the system.

The proposed system is divided into different modules for efficient processing. The modules are as follows:

- Textual Opinions Collection.
- Pre processing & Pos tagging
- Features Identification

- Semantic Database Creation
- Semantic Analysis & Similarity matching
- Opinion Analysis.

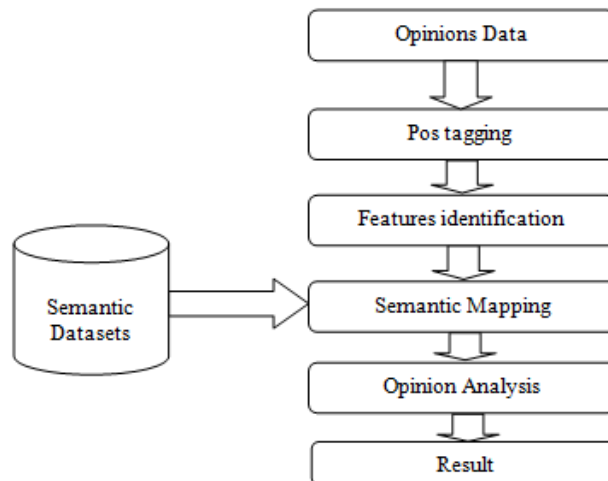


Fig1: Opinions Mining and Summarization System

The Proposed system will be able to solve all or some of the following objectives.

- To provide user with space to write their own opinions as well as taking the opinions from the internet that are available in the form of web blogs, and comments.
- To identify the feature of the organization from the opinions in the pos tagging module.
- To Create the Semantic Datasets for similarity matching
- To find out the features that are relevant to the organization in the semantic mapping module.
- To calculate polarity of the identified features for result formulation in the opinion module.
- Structuring the result for efficient decision making.
- Efficiency of organization can be greatly improved and Solutions can be searched fast for Updation or making improvement.

V. CONCLUSIONS

In this paper, reviews on opinion mining techniques are summarized for mining important information from the freely expressed opinions in the form of textual data. The challenges and scope that are available for analysis of opinions are also stated. The proposed Opinions Mining and Summarization System are described in this paper.

In the future, the work will be carried out in order to implement an efficient system that will be able to solve all or some of the challenges present in the Opinion mining.

REFERENCES

- [1] Katerina Kabassi and Maria Virvou “Combining Decision-Making Theories with a Cognitive Theory for Intelligent Help: A Comparison” IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, VOL. 45, NO. 2, APRIL 2015
- [2] Yingcai Wu, Member, IEEE, Shixia Liu, Senior Member, IEEE, Kai Yan, Mengchen Liu, Fangzhao Wu “OpinionFlow: Visual Analysis of Opinion Diffusion on Social Media” IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 20, NO. 12, DECEMBER 2014
- [3] Desheng Dash Wu, Lijuan Zheng, and David L. Olson “A Decision Support Approach for Online Stock Forum Sentiment Analysis” IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 44, and NO. 8, AUGUST 2014
- [4] Fuji Ren, Senior Member, IEEE, and Ye Wu “Predicting User-Topic Opinions in Twitter with Social and Topical Context” IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 4, NO 4, OCTOBER-DECEMBER 2013
- [5] Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou in their paper “Movie Rating and Review Summarization in Mobile Environment” IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 42, NO. 3, MAY 2012.
- [6] Xiaohui Yu, Member, IEEE, Yang Liu, Member, IEEE, Jimmy Xiangji Huang, Member, IEEE, and Aijun An, Member, IEEE in their paper “Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 4, APRIL 2012.
- [7] Jingbo Zhu, Member, IEEE, Huizhen Wang, Muhua Zhu, Benjamin K. Tsou, Member, IEEE, and Matthew Ma, Senior Member, IEEE in their paper “Aspect-Based Opinion Polling from Customer Reviews” IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 2, NO. 1, JANUARY-MARCH 2011.

- [8] Nathalie Camelin, Associate Member, IEEE, Frederic Bechet, Member, IEEE, Géraldine Damnati, and Renato De Mori, Fellow, IEEE “Detection and Interpretation of Opinion Expressions in Spoken Surveys” IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 18, NO. 2, FEBRUARY 2010.
- [9] Thanh-Son Nguyen, Hady W. Lauw, Member, IEEE, and Panayiotis Tsaparas, Member, IEEE “Review Selection Using Micro-Reviews” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 4, APRIL 2015
- [10] Uros Krcadinac, Philippe Pasquier, Jelena Jovanovic, and Vladan Devedzic “Synesketch: An Open Source Library for Sentence-Based Emotion Recognition” IEEE Transactions on Affective Computing, Vol. 4, No. 3, July-September 2013
- [11] Jelena Jovanovic, Ebrahim Bagheri John Cuzzol Dragan Gasevic, Zoran Jeremic, Reza Bashash “Automated Semantic Tagging of Textual Content” Published by the IEEE Computer Society IT Pro November/December2014 1520-9202/14.
- [12] Chenghua Lin, Yulan He, Richard Everson, Member, IEEE, and Stefan Ruger “Weakly Supervised Joint Sentiment-Topic Detection” from Text IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 6, June 2012.