Survey on Mining Opinions and Sentiments in Social Media

Annie Syrien
Research scholar,
Department of Computer Science
and Applications, Bangalore
University, Bengaluru -56.

M. Hanumanthappa
Professor,
Department of Computer Science
and Applications, Bangalore
University, Bengaluru -56.

S Regina Lourdhu Suganthi
Research scholar,
Department of Computer Science
and Applications, Bangalore
University, Bengaluru -56.

Abstract:- The development of web2.0 has envisaged great impact in processing user generated content into useful resources through various mining algorithms. The focus has been placed on mining social media for opinions, sentiments, attitudes and emotions. In the present days the market largely depends on promoting their products through social media and hence it has become an integral part of business. The promotions can be for the products, events, individuals, topics or politics and the like. Though promotions have very important advantage in gaining high profit for products, new business launch, unpredicted fame for individuals and so on, where the desired market aspirers go to an extent of investing a huge corpus to achieve, may not lead to success always. Thus the other alternative is social media, where in the user generated content is genuine, due to the fact that the user is free from external influences and just gives the real information. This information acts as unbiased source to be processed using mining techniques to derive meaningful patterns to improve business activities. In this survey paper various techniques for sentiment analysis and opinion mining of social media data has been discussed and also reviewed the different works that has been carried out in this field.

Keywords— Social media, opinion mining, sentiment analysis

1. INTRODUCTION

The data available online are analyzed in order to extract useful information in various fields such as business, politics, entertainment, organization, events and topics. Advancement in hardware devices, anytime anywhere access to information has made larger impact on the way one can communicate with other and share the resources. People´s opinions have become a key factor for making decisions, not only for individuals but also for government and commercial sectors [7].

In social media millions of people share their ideas, views, reviews, ratings on various events, topics, organizations, business products where these information can be further analyzed in order to extract the useful information where it can be used to draw the useful conclusions and suggestions for future implementations and developments. In the recent past opinion mining and sentiment analysis gained a vast scope in the field of web data mining. The fast growth of online user generated contents in blogs, wikis and web forums lead to the era of processing to derive important information. Due to the huge data production from the online social networking users,

the opinion mining and sentiment analysis is attracting large set of people both from industry and academia. Jinhui et al. [4] states that the rapidly evolving social networks provide a platform for communication, information sharing, and collaboration among friends, colleagues, alumnus, business partners, and many other social relations. To be accompanied by, increasingly rich and massive heterogeneous media data have been generated by the users, such as images, videos, audios, tweets, tags, categories, titles, geo-locations, comments, and viewer ratings, which offer an unprecedented opportunity for studying novel theories and technologies for social media analysis and mining. Recently more and more research is been carried out in this field.

1.1 SOCIAL MEDIA

Social media is an internet based communication tool that empowers people to share information. To understand better the term social media, social indicates associating with people and spending time in order to develop their relationships whereas media indicates tool for communication such as internet, TV, radio, newspaper etc. here our focus is internet. So social media can be stated as, An electronic platform for socializing people. Some example of social media sites are Facebook, Twitter, Youtube, LinkedIn, Digg etc. Initially people involved in social media to associate with their friends and lost friends, gradually they improved to the status of updating and consuming any information on social media, these led to vast generation of user data which could be further processed for future development.

1.2 OPINION MINING AND SENTIMENT ANALYSIS

Opinions are subjective expressions of emotions, feelings, attitudes or sentiments towards entities, events and their properties. One important aspect of opinions is the fact that they have targets: opinions are expressed for objects (i.e. entities or events) and their attributes [5].

Opinion mining is a field of study where in a large scale opinions are studied in order to classify positive or negative sentiments through sentence level, document level and feature level opinions. There are three components of opinion mining, the first one is opinion holder, the second is object and the third is opinion. The opinion holder is a person or an organization that holds opinion on specific object. The object is an item on which opinion is expressed such as topic, organization, event, individuals and product. The opinion

indicates the sentiments, emotions, such as positive, negative and neutral.

Opinion Mining and Sentiment Analysis has been attracting many researchers from the past few years due the evolution of Web 2.0. It has been proven by many researchers who conduct survey or analyze the data that has been obtained from the social media prove to be truth due to its self-substantial nature of the data, the reason is users in the social networking sites quote anything without any obligations from others. Due to its reliable nature of data the academia and industry focus on social media data for their business improvements, enhancements, to learn the present scenario, attitudes, emotions regarding any objects which is their focus. There been a lot of research in order to mine the data through efficient algorithms to improvise its speed and proficiency of the obtained information.

The sentiment analysis itself a field of study which can be further classified into three levels such as

- Document level sentiment analysis
- Sentence level sentiment analysis
- aspect level sentiment analysis

Whereas the aspect level is further classified as aspect and opinion extraction, sentiment lexicon analysis, opinion summarization. The aspect and opinion extraction is subdivided as entity extraction and categorization, aspect extraction and categorization, opinion extraction and categorization, time extraction and standardization. The sentiment lexicon analysis compartmentalized as Senti-WordNet, WordNet-Affect, SenticNet. Finally the Opinion summarization is sub classified as quantitative and qualitative [1].

1.3 METHODS FOR SENTIMENT ANALYSIS

(a) Sentiment Classification Using Supervised Learning
Sentiment classification is divided as a two-class classification problem, positive and negative. Training and testing data used are normally done through product reviews. Since online reviews have rating scores assigned by their reviewers, e.g., 1-5 stars, the positive and negative classes are determined using the ratings [9]. For example, the review with a range from 5 to 4 can be considered as positive review whereas 2 to 1 as negative review. The review which is in middle can be rated as neutral. Most research papers do not use the neutral class to make the classification problem easier.

(b) Sentiment Classification Using Unsupervised Learning
The unsupervised classification takes place based on the separation parameters that is chosen to specify the difference. The objective is to group multiband spectral response patterns into clusters that are statistically separable.
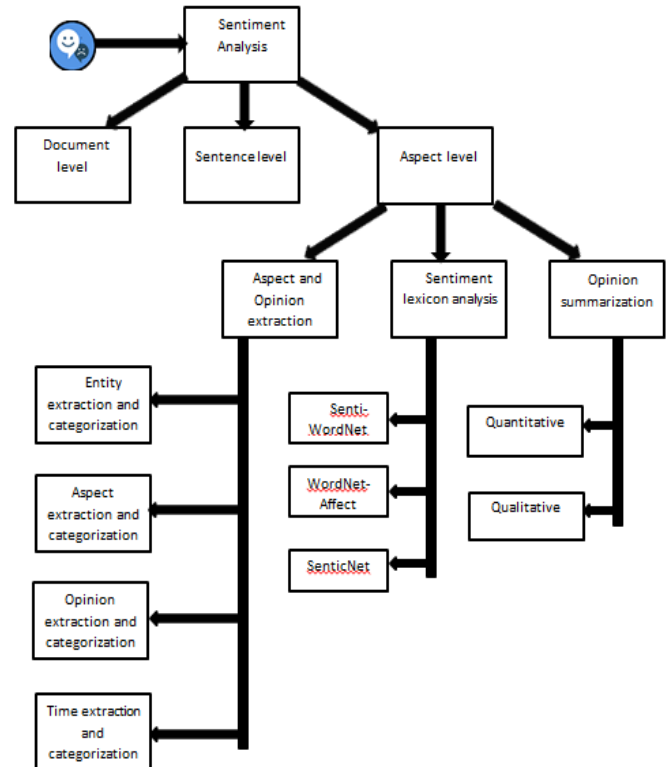

FIGURE 1 : SENTIMENT ANALYSIS

II LITERATURE SURVEY

As Opinion mining and sentiment analysis is a main focus for researchers, there been a many research works are carried out which are listed below.

Togir et al. [1] overviews the various aspect extraction techniques and approaches and describes classification levels of sentiment analysis, aspect extraction approaches, methodologies such as explicit aspect extraction, where they demonstrate the table of summaries such as summary of unsupervised techniques frequency based FB2, bootstrapping BST1, rule based algorithms supports implicit aspects. In the summary of semi-supervised techniques bootstrapping BST2, double propagation DP, DP1, page rank and lexicon based EXPRS algorithms put up implicit aspects. In supervised techniques no algorithms supports implicit aspects. The conclusion is most of the research has focused on unsupervised and semi-supervised learning for implicit aspects while supervised approach is not much explored due to time consumption of training the dataset.

Anwar Hridoy et al. [2] demonstrated the methodology to determine the opinion or sentiment of a product at different locations across male or female users on Twitter. For their analysis they chose iPhone 6 as their research product reason being availability of sufficient data. The data was collected through Twitter public API which allowed developers to extract data programmatically. The collected data was filtered due to the casual nature of tweets. NamSor tool was used classify the gender of each tweet. They carried out their work by various stages such as data extraction, data preprocessing, and implementation phases. The comparisons of graphically

presented data were done with respect to the real world scenarios to find the perfection of methodology.

Yan Li et al. [3] proposed a novel model in order to bring realization on aspect based sentiment summarization in an integrative way; that is feature extraction and clustering composed through the system; collocation orientation disambiguation and sentence sentiment strength calculation whereas pattern based bootstrapping used in collocations of product features and opinion words extraction. The two important measurements are prevalence and reliability is used to exploit patterns and features. The obtained features are clustered into aspects. Every cluster is allotted a weight based on arithmetic means of feature similarities and confidences. The dynamic sentiment ambiguous adjectives are determined inside opinion collocations. Ultimately, sentiment strengths of opinion for each aspects are computed according to a set of fine grained and stratified scoring formulae.

Jonathan et al. [6] surveyed on significant annotations about subjectivity in linguistics and psychology, in order to enlarge the standard opinion analysis perspective followed by computational sentiment analysis which existed long ago. Putting together these two slants leads to assessing the gap between the broader notation of subjectivity analysis and the subfields that language technology research focuses on.

Mikalai et al. [8] reviews the evolution of sentiment analysis and opinion mining from past few years, discusses the development of development of new research areas such as contradiction analysis. They also overview various research directions such as subjectivity analysis, sentiment polarity, dictionary approach, opinion aggregation, opinion quality and spam.

Manolis et al. [10] retrospected a variety of opinion mining methods for examining product opinions and reviews put up on web. For analyzing citizen's contributions in public policy, the review plays an important role. Textual citizen contributions in public policy debates, both for assessing contributors' general attitudes – sentiments for decision under discussion and also for extracting the main issues such as improvement suggestions, implementation barriers, negative and positive aspects. The other is corresponding attitude sentiments. A basic framework for the use of opinion mining methods in eParticipation has been articulated.

Alexandar et al[11], in their research work talks about augmentation structures. As the sentiment analysis is focused study from the point of research many researchers collect the data and analyze it through the number of alike words largely ignore semantic and structural aspects of content, whereas Alexandar et al proposes a model sentiments with regard to economic is analyzed through augmentation structures.

III SUMMARY OF LITERATURE REVIEW

Enormous research work is carried out in the opinion mining and sentiment analysis. The table 1 summarizes the different features and techniques used in the field of opinion mining and sentiment analysis.

Table 1: Summary of Literature Review

Sl. No.	Author and Year	Features and Techniques
1	Togir et al., 2016	Aspect extraction techniques, frequency based FB2, bootstrapping BST1, rule based algorithms, double propagation DP, DP1, page rank and lexicon based EXPRS algorithms.
2	Anwar Hridoy et al., 2015	IPhone6 Twitter review, Twitter public API, . NamSor tool.
3	Yan Li et al., 2014	Aspect based sentiment summarization, arithmetic means of feature similarities.
4	Jonathan et al., 2014	Significant annotations, computational sentiment analysis.
5	Mikalai Tsytsarau et al., 2012	Contradiction analysis, subjectivity analysis, sentiment polarity, dictionary approach, opinion aggregation, opinion quality and spam.
6	Manolis et al., 2011	Feature-Based Sentiment Analysis, Ontology-Based Sentiment Analysis, analyzing citizens contributions in public policy.
7	Alexander et al, 2010	Augmentation structures.

IV CONCLUSION

From the discussions above its clear that many researchers focus on developing opinion mining and sentiment analysis algorithms in order to improvise them and prove their efficiency for both supervised and unsupervised sentiment classification. In the research the social media, opinion mining and sentiment analysis, methods for sentiment analysis is been discussed, followed by few related work on the aspect extraction in sentiment analysis, tweeter opinion mining on iphone6 review, mining product aspects in an integrative way, computational sentiment analysis, contradiction analysis, analyzing citizens contributions in public policy, sentiment analysis through augmentation structures. From the research which is conducted it is evident that opinion mining and sentiment analysis used supervised learning techniques in order to classify their data. Unsupervised and semi-supervised learning techniques are useful when the implicit aspect extraction techniques are used and most suitable.

REFERENCES

[1]. Togir et al, "Aspect extraction in sentiment analysis: comparative analysis and survey", Springer Science 2016, Artif Intell Rev DOI 10.1007/s10462-016-9472-z

[2]. Anwar Hridoy et al, "Localized twitter opinion mining using Sentiment Analysis", Decision Analytics, (2015)2:8 DOI 10.1186/s40165-015-0016-4

[3]. Yan Li et al, "A holistic model of mining product aspects and associated sentiments from online reviews", Springer Science (2014) DOI 10.1007/s11042-014-2158-0

[4]. Jinhui Tang et al, "Social media mining and knowledge discovery", Springer, Multimedia Systems (2014) 20:633–634 DOI 10.1007/s00530-014-0423-8

[5]. Georgios Petasis et al, "Sentiment Analysis for Reputation Management: Mining the Greek Web" Springer International Publishing Switzerland (2014) LNAI 8445, pp. 327–340

[6]. Jonathan et al, "Sentiment Analysis: What's Your Opinion?", Springer International Publishing Switzerland (2014), C. Biemann, A. Mehler

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCRIT - 2016 Conference Proceedings

(eds.), Text Mining, Theory and Applications of Natural Language Processing, DOI 10.1007/978-3-319-12655-5_9

[7]. Mohammed Almashraee et al, "Feature Extraction Based on Semantic Sentiment Analysis", Springer-Verlag Berlin Heidelberg (2013), BIS Workshop, LNBIP 160, pp. 270–277

[8]. Mikalai Tsytsarau et al, "Survey on mining subjective data on the web", Springer, Data Min Knowl Disc (2012) 24:478–514 DOI 10.1007/s10618-011-0238-6

[9]. Bing Liu, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, May 2012

[10]. Manolis et al, "A Review of Opinion Mining Methods for Analyzing Citizens' Contributions in Public Policy Debate", International Federation for Information Processing, ePart (2011), LNCS 6847, pp. 298–313

[11]. Alexander et al, "Mining Economic Sentiment Using Argumentation Structures", Springer-Verlag Berlin Heidelberg (2010), ER Workshops, LNCS 6413, pp. 200–209.