



ISSN: 2350-0328

## International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 7 , July 2017

# Survey on Social Media Analysis

T. Miranda Lakshmi, A. Sophia Vilsana, Dr. V. Prasanna Venkatesan

Research Scholar, Department of Computer Science, Bharathiyar University, Coimbatore, India.  
Research Scholar, Department of Computer Science, St. Joseph's College of Arts and Science, Cuddalore, India.  
Assoc. Prof., Department of Banking Technology, Pondicherry University, Pondicherry, India.

**ABSTRACT:** Social networking sites is an online communicating platform, large number of user to exchange their own ideas and opinions on that platform for collecting, tracking and analyzing public sentiment. Sentiment analysis can give some critical information for decision making process in different domains. In this paper presents the literature survey for referencing needs of researchers. The literature survey is based on data collected from different paper.

**KEYWORDS :** Sentiment Analysis, Face book, Twitter, Social Networking.

### I. INTRODUCTION

Social networking technologies, such as those provided by online Social networking sites such as facebook (where users have an online social “friendship” relationship) and twitter as a microblogging website (where the online social relationship can be uni-directional or reciprocal), are attract and encourage millions of users to share and exchange their ideas and opinions and to participate in events. Every day, social media users express their feelings towards various subjects such as politicians, sports men, governmental organizations, entertainments groups and consumer products (Liu, 2012), create a large amount of data. One of the most popular social networks is Facebook, with a reported 1.2 billion active users monthly and 725 million active users daily (Facebook 2013) and Facebook users upload 350 million new images every day (Henschen, 2013). Social media is achieving more and more importance as a channel for gathering information, simulated new ways of interactions, shaping new forms in which people communicate, make decisions, socialize, collaborate with others, learn, entertain themselves, interact with each other or even do their shopping (Hanna et al., 2011). This paper have the collection of sentiment analysis reviews for referencing need of researcher in this topics.

### II. LITERATURE REVIEW

Social media involved in various fields sports, political, entertainment, educations, health care, agriculture, government and etc. Sentiment analysis has been used to determine emotional difference between genders on Myspace (Thelwall et al., 2010), and study levels of positive and negative sentiments in Facebook (Thelwall et al., 2010) and Twitter comments (Bollen et al., 2011) sentiment analysis has also been used to predict election outcomes (Tumasjan et al., 2010), and it was demonstrated that sentiment relating to new movie releases, combined with tweet frequency, was more accurate at predicting box office revenue than the Hollywood stock market (Asur & Huberman, 2010). Social media like facebook to explain fan motivation and relationship (Constantino et al., 2013) and providing a more complete picture of the nature of fan motives (Wann et al., 2001) will better enable sport teams to differentiate their brands (Green, 2001) and to maximize their use of online channels of interaction (Seo & Green, 2008), especially social media (Earnhardt et al., 2012). (Seo & Green, 2008) developed a fan motivation scale, relating to professional sports teams' websites, consisting of ten dimensions: fan ship, interpersonal communication, technical knowledge, fan expression, entertainment, economic, pass time, information, escape, and support. These authors acknowledge that the Internet more broadly could satisfy a range of needs in excess of traditional media given its interactive capabilities.

Twitter data has already been used by different people to predict stock market prediction (Boollen et al., 2011), identify the client with negative sentiments (Thet et al., 2010). Facebook communities' data has been used to discuss health care related to health screening, smoking, alcohol consumption, drug use and to predict patient's emotional states such as mood, happiness, depression (Fowler & Christakis, 2008). The main aim of sentiment analysis



ISSN: 2350-0328

# International Journal of Advanced Research in Science, Engineering and Technology

Vol. 4, Issue 7 , July 2017

is to determine the attitude of users on a particular topic and the purpose of sentiment analysis to offer a significant insight into online communication.

## A. Social Media

Social media is a collection of online communication channels can be used to sharing personal, official and other information and provide interaction and collaboration to their family, friends, communities and others. Forums, blogs, social networking sites, social bookmarking, wiki and social curation are the different types of social media (Barbie & Liu, 2011).

## B. Social Media Analytics

Social media analytics that can be used to gather the information from social media sites like blogs, social networking sites (Facebook, Twitter, Wikipedia, google+ and pinterest) for analyzing these data to make good decision for business. One of the most wieldy using social media analysis technique is sentiment analysis.

## C. Sentiment Analysis

Sentiment analysis or opinion mining is a discipline which deals with analyzing and classifying subjective opinions, sentiments and emotions of people towards products, organizations, individuals and other topics (Feldman, 2013) expressed in text, such as tweets, review, forums, blogs, and news. Sentiment analysis of social media data has received a lot of research attention in recent years, largely due to the increasing use of social networking technologies that enable citizen to self-report their opinions as frequently as they wish on a wide range of topics. The method of sentiment analysis generally requires the identification of an entity to which the opinion is focused on (a person); attributes of the entity (the person political perspective); views, attitudes or feelings towards the entity and its attributes (commonly defined as sentiment); an opinion holder (comments posted in response to a news story); and a time at which the sentiment was expressed (Liu, 2012). The textual opinion mining has been to classify sentiment using a pre-defined set of classers (positive, negative or neutral). Social networking sites have two most important topics, they are community detection and sentiment analysis. Community detection to find groups of associated individuals within network and sentiment analysis attempts how individuals are feelings

Sentiment analysis work can be divided into six tasks which are

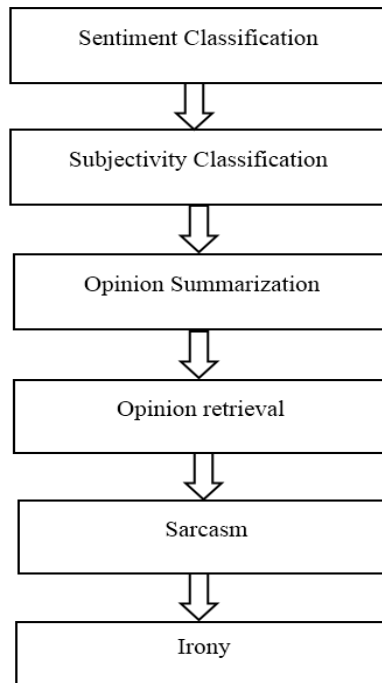


Fig1. Sentiment Analysis work

#### **D.Sentiment Analysis Methods**

Lexicon-based methods, machine learning-based methods, and hybrid methods are the three common approaches for sentiment analysis (Medhat et al., 2014).

##### **1)Lexicon-Based Method**

A lexicon-based method uses a sentiment lexicon which includes sentiment words and phrases with assigned numeric scores. These scores reveal if sentiment phrases are positive, negative or neutral, their intensity and their emotional orientations.

##### **2)Machine Learning- Based Method**

A machine learning-based method uses machine learning algorithms such as Naïve Bayes, Support Vector Machine, Regression and Neural Network.

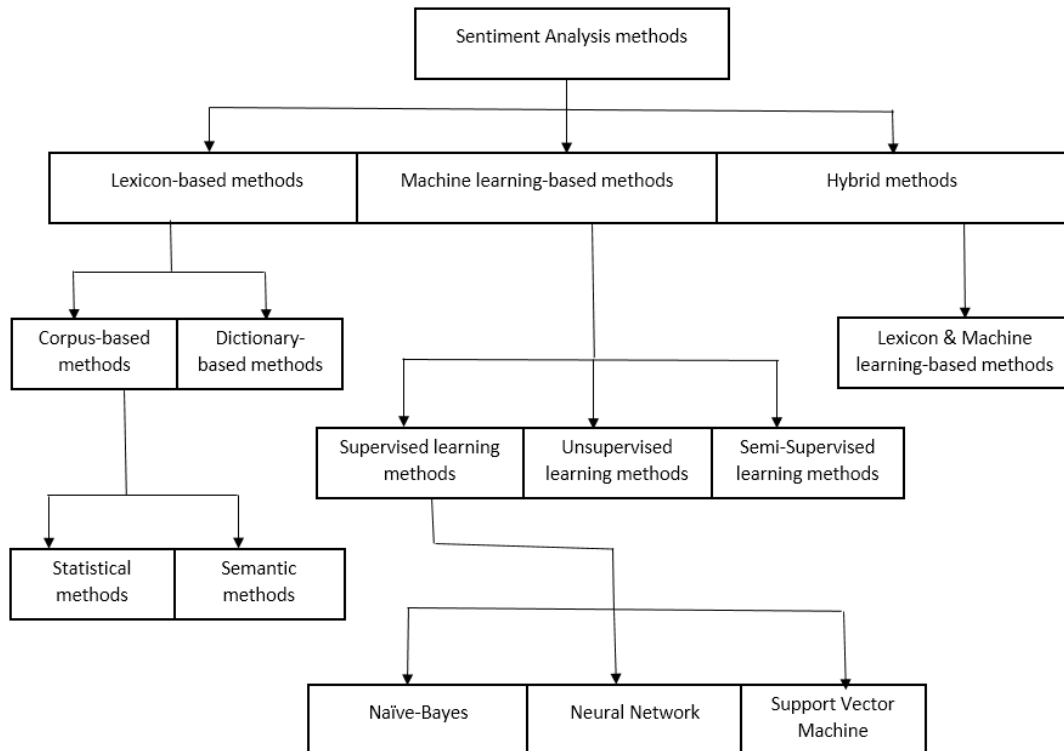


Fig2. Sentiment Analysis Methods.

### 3)Hybrid Method

A hybrid method is a combination of lexicon-based and machine learning-based methods.

Machine learning has shown that probabilistic machine classifiers such as Naive Bayes (NB) and support vector machines (SVMs) stand out as having produced the best result for sentiment analysis (Park & Paroubek, 2010), (Go, 2009). The machine classifiers gives decisions based on the learned probability and unobserved text can be added to pre-defined categories based on the features it contains. The classification decision can be made by using Bayes Theorem, which frames the probability of a new data belonging to a particular class given previous knowledge of data items with a known classification. The main weakness of this approach that the hidden data is not learned during the training period, so it is not helpful in classification.

Support Vector Machine (SVM) to classify unseen data by increasing the distance between clusters, clusters can be created using training data and finding an optimal solution and it is particularly used for text classification (Thelwall et al., 2010). Sentiment analysis features typically include set of single words, double words or pairs and triple words found in text (Thelwall et al., 2011). Several research reports gives best performance for unigrams (single word) (Lee et al., 2002) other reports that bigrams and trigrams (two and three word combinations) performs better (Dave et al., 2003).Lexicon based method also have to produced good result. Corpus method can be used to analyse the product reviews and dictionary method used to analyse Amazon's mechanical Turk data (Taboada et al., 2011). Sentiment lexicon that can help sentiment identification in semantic oriented sentiment classification (Hu and Liu, 2004). Some sentiment lexicons have been developed such as (Baccianella et al., 2010) and (Esuli&Sebastiani, 2006) proposed SentiWordNet 1.0 and 3.0 to defined three sentiment labels, (Cambria et al., 2016) proposed SenticNet 2.0,3.0 and 4.0 to define sentiment polarities and value for words and (Thelwall et al., 2012) proposed SentiStrength to define sentiments used in social networks.

**III. METHODOLOGY****A. Problem Statement**

In order to analyse the popularity of people in various fields such as sports, entertainments, cultural, political and educational. It is very easy to give awareness about any new diseases, vaccines, products, norms to people because a person is well-known by people.

**B. Experimental Design****1) Data Collection**

Data used in this paper is a set of facebook metrics data, popularity of news data and blogs data from UCI repository. In total, over 39645 posts in which the posts belong to different categories: sports, entertainment, political, spatial and etc. Each post includes the following information: post id, time, number of images, number of videos, shares etc. These data can be efficiently handled by R studio tool and big data based algorithms that could possibly handle huge amount of data in quick time.

**IV. CONCLUSION**

Sentiment analysis on social media is a one of the most important research topic. This paper presented an overview of sentiment analysis on social networks and different methods on sentiment analysis. Through this literature survey the related work can be done for referencing needs of researchers. And this paper give data collection and proposal of popularity prediction. This data collection will also help to analyse the popularity of a person in social media. This research will used to improve a person popularity level and some products, awareness will easily reach the people.

**REFERENCES**

- [1] Go, L. Huang, R. Bhayani, Twitter sentiment analysis, Final Projects from CS224N for Spring 2008/2009, Stanford Natural Language Processing Group, 2009.
- [2] A.Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, Seventh Conference on International Language Resources and Evaluation, 2010.
- [3] Asur, S., Huberman, B., et al. (2010). Predicting the future with social media. In International conference on web intelligence and intelligent agent technology (WI-IAT): 1(pp. 4 92–4 99). IEEE.
- [4] Barbier, G., & Liu, H., Data mining in social media. In C. C. Aggarwal (Ed.), Social network data analytics (pp. 327–352). United States: Springer 2011.
- [5] Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool, 2012.
- [6] Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, Empirical Methods in Natural Language Processing, 2002.
- [7] B.Liu, sentiment analysis: A multi-faceted problem, IEEE Intelligent Systems 25.3 (2010): 76-80.
- [8] Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of international conference on language resources and evaluation (pp. 2200–2204).
- [9] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2, 1–8.
- [10] Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2016). SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In Proceedings of 26th international conference on computational linguistics.
- [11] Fowler, J. H., & Christakis, N. (2008). Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham heart study. British Medical Journal, 337, a2338.
- [12] Henschen, Doug (2013). Facebook on Big Data Analytics: An Insider's View. Information week. Retrieved May 7, 2013.
- [13] J. Ahktar, S. Soria, Sentiment Analysis: Facebook Status Messages, Stanford University, 2009.
- [14] J. Bollen, B. Gonvalves, G. Ruan, H. Mao, Happiness is assortative in online social networks, Artif. Life 17 (2011) 237–251.
- [15] K. Dave, S. Lawrence, D. Pennock, Mining the peanut gallery: opinion extraction and semantic classification of product reviews, 12th International conference on WorldWideWeb, ACM, New York, NY, USA, 2003.
- [16] M. Thelwall, K. Buckley, G. Paltogou, D. Cai, A. Kappas, Sentiment strength detection in short informal text, J. Am. Soc. Inf. Sci. Technol. 61 (2010) 25442558.
- [17] M. Thelwall, K. Buckley, G. Paltogou, Sentiment in Twitter events, J. Am. Soc. Inf. Sci. Technol. 62 (2011) 406–418.
- [18] M. Thelwall, K. Buckley, G. Paltogou, Sentiment in Twitter events, J. Am. Soc. Inf. Sci. Technol. 62 (2011) 406–418.
- [19] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5 (4), 1093–1113.
- [20] R. Feldman, Techniques and applications for sentiment analysis, Communications of the ACM 56.4 (2013): 82-89.



ISSN: 2350-0328

**International Journal of Advanced Research in Science,  
Engineering and Technology**

**Vol. 4, Issue 7 , July 2017**

- [21] S. Asur, B.A. Huberman, Predicting the Future With Social Media, 2010.
- [22] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., &Stede, M..“Lexiconbased methods for sentiment analysis”. Computational linguistics, 2011:37(2), 267-307.
- [23] Thelwall, M., Buckley, K. ,&Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology,63 (1), 163–173.
- [24] Thet, T. T., Na, J.-C. , &Khoo, C. S. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of Information Science.
- [25] Tumasjan, T. Sprenger, P. Sandner, I. Welp, Predicting elections with Twitter: what 140 characters reveal about political sentiment, International AAAI Conference on Weblogs and Social Media, Washington, D.C., 2010.