# CURRENT TRENDS OF OPINION MINING AND SENTIMENT ANALYSIS **IN SOCIAL NETWORKS**

# Sudipta Roy<sup>1</sup>, Sourish Dhar<sup>2</sup>, Arnab Paul<sup>3</sup>, Saprativa Bhattacharjee<sup>4</sup>, Anirban Das<sup>5</sup>, Deepjyoti **Choudhurv<sup>6</sup>**

<sup>1</sup>Department of Information Technology, Triguna Sen School of Technology, Assam University, Silchar <sup>2</sup>Department of Information Technology, Triguna Sen School of Technology, Assam University, Silchar <sup>3</sup>Department of Information Technology, Triguna Sen School of Technology, Assam University, Silchar <sup>4</sup>Department of Information Technology, Triguna Sen School of Technology, Assam University, Silchar <sup>5</sup>Department of Information Technology, Triguna Sen School of Technology, Assam University, Silchar <sup>6</sup>Department of Information Technology, Triguna Sen School of Technology, Assam University, Silchar

### Abstract

Sentiment analysis is currently treated as a famous topic as well as in the area of research with significant applications in both industry and academia. This paper represents the importance and applications of opinion mining and sentiment analysis in social networks. We have described the basic concepts, challenges and comprehensive study in different sections in this paper. This paper reflects as an overview of opinion mining and sentiment analysis in social networks.

\*\*\*\_\_\_\_\_

Keywords: Social Media, Opinion Mining, Sentiment Analysis, Social Network.

# **1. INTRODUCTION**

Graph representation is the easy way to represent a social network. We can also represent a social network as in tabular form [1]. The main goal of analysis a social network is to study the structural properties of a network. Structural properties of a network investigate the global properties of the individual nodes in the network. Many individual nodes form the groups (communities) with dense intra-connection among the nodes in a social network. Centrality is the important measurement of a social network which determines the structural importance of a node in a network. There are three types of centrality: degree centrality, betweenness centrality and closeness centrality. Prestige is the number of in-links to a node in a network.

From the point of view of opinion mining the ability to assess the node's prestige is essential as it allows differentiating between opinions of different individuals. More specifically, node's prestige allows assigning different weights to opinions and associating more importance to opinions expressed by prominent individuals. The identification of influential individuals is another factor that is often considered in opinion mining. An influential individual node does not have to be necessarily characterized with high degree centrality to influence the average opinion within the network. Generally, such individual nodes are characterized by high betweenness centrality.

Communities are densely inter-connected and have more less connections to the nodes of a community from any other communities. The opinion mining algorithm can be used for the opinion prediction when the information is provided on group membership of an individual node. Opinion mining is the domain of natural language processing and text analytics that aims at the discovery and extraction of subjective qualities from textual sources. Opinion mining tasks can be generally classified into three types. The first task is referred to as sentiment analysis and aims at the establishment of the polarity of the given source text (e.g., distinguishing between negative, neutral and positive opinions). The second task consists in identifying the degree of objectivity and subjectivity of a text (i.e., the identification of factual data as opposed to opinions). This task is sometimes referred to as opinion extraction. The third task is aims at the discovery and summarization of explicit opinions on selected features of the assessed product. Some authors refer to this task as sentiment analysis. All three classes of opinion mining tasks can greatly benefit from additional data that may be provided from the social network.

Monitoring social media to spot emerging issues and trends and to assess public opinion concerning topics and events is of considerable interest to security professionals; however, performing such analysis is technically challenging. The opinions of individuals and groups are typically expressed as informal communications and are buried in the vast, and largely irrelevant, output of millions of bloggers and other online content producers. Consequently, effectively exploiting these

data requires the development of new, automated methods of analysis [2][3].

In this paper, we have discussed the basic concepts on current techniques of opinion mining and social networks in the second section. Then we have described the challenges of opinion mining imposed by social media in the third section. In the fourth section we have shown the comprehensive study. Then we conclude our paper.

# 2. BASIC CONCEPTS

### 2.1 Current Opinion Mining Techniques

The identification of the opinion polarities and strength inside a text is a fairly recent area of study with plenty of applications and challenges. Hatzivassiloglou and Mckeown [4] are among the first to deal with opinion classification. They focus on adjectives and they study phrases where adjectives are connected with conjunction words such as 'and' or 'but'. They construct a log-linear regression model so as to clarify whether two adjectives have the same orientation. The accuracy of this task is declared to be 82%. Their technique is described in the following steps:

- (a) Extract from a text the conjoined adjectives that are connected with the words 'and'/'but' etc.,
- (b) Run a supervised algorithm that builds a graph where the nodes are the adjectives and the links determine same or opposite orientation,
- (c) Run a clustering algorithm to separate the graph into two classes,
- (d) Assume that the cluster with the highest frequency is the one that shows positive orientation.

In 2003, Turney and Littman [5] use a point-wise mutual information (PMI) and a latent semantic analysis (LSA) measure to find out the statistical relation between a specific word and a set of positive or negative words. They construct a seed set which contains words that can be classified as either positive or negative independently of the context e.g. 'best' is always positive. The LSA-based measure gives better results than the PMI-based one. Wiebe [6] deals with the distinction between objective and subjective sentences i.e. between facts and opinions. They deal with 3 subjectivity types: positive, negative and speculation.

To construct a seed set with the right adjectives is not a straightforward task. In 2008, Harb et al. [7] present a work in which they generate automatically a dictionary of adjectives. Initially they collect their data by getting web documents that contain negative and positive opinions.

In the field of Opinion Mining, a significant work is carried out by Hu and Liu [8] in 2004. They deal with product reviews written by customers on web sites. Their objective is to produce a structured summary that informs about positive or negative statements that are made for product features.

### 2.2 Current Social Network Techniques

The Social Network Analysis deals with the analysis of the relationships that exist between entities in a social network. For instance, in a social network of people, the analysis can include who is friend with whom, who can influence which group of people, whom can have access to the information that goes through this network etc.

In 2006, Fisher et al. [9] analyze newsgroups by applying Social Network techniques and they interpret online communities by assigning roles to the members of the groups. This is done by observing how people relate to each other in a graph-based model of post-reply relations.

Java et al. [10] analyze the Twitter's social network and the intentions of the associated users in order to understand the reason why people use such networks. They identify the communities that are formed; they categorize them into communities that create information, communities that receive information and communities that exist only because of friendship.

In 2007, Scripps et al. [11] introduce a new measure that defines the number of communities to which a node is attached. Using this measure they assign roles to nodes by considering the community structure in the network of the node.

Maurel et al. [12] have analyzed forums in the domain of tourism and they have extracted information regarding user sentiments and tourist destinations. They apply syntactic and semantic processing techniques and they adapt the grammar rules or the opinion words they try to identify according to the domain.

#### 3. CHALLENGES IMPOSED BY SOCIAL MEDIA

There are some challenges in opinion mining imposed by social media we have discussed in below:

#### **3.1 Relevance**

Even when a crawler is restricted to specific topics and correctly identifies relevant pages, this does not mean that every comment on such pages will also be relevant. This is a particular problem for social media, where discussions and comment threads can rapidly diverge into unrelated topics, as opposed to product reviews which rarely stray from the topic at hand.

# **3.2 Target Identification**

One problem faced by many search-based approaches to sentiment analysis is that the topic of the retrieved document is not necessarily the object of the sentiment held therein. One way in which we deal with this problem is by using an entitycentric approach, whereby we first identify the relevant entity and then look for opinions semantically related to this entity, rather than just trying to decide what the sentiment is without reference to a target, as many machine learning approaches take.

#### 3.3 Negation

The simpler bag-of-words sentiment classifiers have the weakness that they do not handle negation well; the difference between the phrases "not good" and "good" is somewhat ignored in a unigram model, though they carry completely different meanings. A possible solution is to incorporate longer range features such as higher order dependency structures.

# **3.4 Contextual Information**

Social media, and in particular tweets, typically assume a much higher level of contextual and world knowledge by the reader than more formal texts. This information can be very difficult to acquire automatically. For example, in 2011, one tweet in the political dataset used in [13] likened a politician to Voldemort, a fictional character from the Harry Potter series of books. One advantage of tweets, in particular, is that they have a vast amount of metadata associated with them which can be useful, not just for opinion summarization and aggregation over a large number of tweets, but also for disambiguation and for training purposes. Examples of this metadata include the date and time, the number of followers of the person tweeting, the person's location and even their profile.

# 3.5 Volatility Over Time

Social media, especially Twitter, exhibits a very strong temporal dynamic. More specifically, opinions can change radically over time, from positive to negative and vice versa. The different types of possible opinions are associated as ontological properties with the classes describing entities, facts and events, discovered through information extraction techniques similar to those, and semantic annotation techniques similar to those in [14] which aimed at managing the evolution of entities over time.

#### **3.6 Opinion Aggregation**

In classical information extraction, aggregation can be applied to the extracted information in a straightforward way: data can be merged if there are no inconsistencies, e.g. on the properties of an entity. Opinions behave differently here, however: multiple opinions can be attached to an entity and need to be modeled separately, for which we advocate populating a knowledge base.

# 4. COMPREHENSIVE STUDY

The comparative analysis of the existing system is summarized [15] in the Table 1 with their appreciations and limitations.

Table 1	l
---------	---

<u></u>			
SI no.	METHODOLO GY	APPRECIATIO N	LIMITATION S
1	Feature based opinion mining	Product features and product components are analyzed separately. Accuracy of feature mining is high.	Method to handle neutral opinion has not been concentrated.
2	Subjective and objective based opinion mining	Spelling errors and tagging of reviews is done effectively	Supervised learning has to be used for further improvement in accuracy
3	Affectivity based opinion mining	Collaborative approach to support various web mining communities has been dealt.	This method classifies each review as taxonomy of elements and affectivity of element and it does not perform the process of disambiguation prior to this step.
4	Semantic role labeling and polarity computation	Effective method of feature lexicon construction has been made.	Feature polarity is divided only to five levels which will not be sufficient for a large corpus of reviews
5	Feature based approach involving syntax tree pruning and kernel based classifiers	Pruning strategy is explicitly concentrated for adjectives which improve the performance. Kernel classifiers improve sentence level classification of reviews.	Pruning of syntax tree was essentially developed to remove unwanted elements in review but sometimes this may remove the most essential information required for the

			classification
6	Polarity based feature extraction method	Polarity is estimated not based on just the nature of objective but also based on the context in which the objective is used	of reviewsChangeinpolarity due topresenceofadjectivesatvariouspositionsofsentenceisanalyzed
7	Lingual knowledge based opinion mining	Reviews are not limited to one particular domain or site. Reviews from different sites are collected using crawlers and the variation in the formats of the posts is also handled	Approaches used to classify the sentiment have not been deeply discussed.
8	Feature based approach using machine learning method	Three kinds of feature sets are covered, bag of words, appraisal phrases and n gram features. This method has been tested using svm, naïve bayes and decision tree classifiers.	This approach ignores the location of the polarity words. N- gram feature set does require exact order of processing as a result certain interesting patterns in mining may be missed.
9	Feature driven opinion summarization method	Both product independent features and product dependent features are analyzed. Concept-net and word-net are used to extract the feature sets	Handling of neutral reviews is not performed. SVM classifier is dependent on the scores for anchor features
10	Semantic orientation based opinion mining	In this method the usage of a word, phrasal form of a word and the syntactic class to which it belongs to is identified using how-net. Also the	Requires a lot of manual processing of reviews to train the system and conclude the semantic orientations.

11	Feature based opinion mining	sentiment orientation of each word is identified by considering potency, evaluative and activity factors. Hence the chance of mis-predicting the orientation is too low. Co-occurrences of words are considered in analysis which increases the weightage for a particular feature. Frequency is calculated for each of the terms in the reviews and a candidate list of features is created. This reduces the cost of missing a feature. Pattern	Reviews are categorized as relevant and irrelevant on the basis of the domain they commented on. But this classification has been done on the basis of words present in the review.
		match is done for classifying the reviews.	
12	Maximum entropy model for product feature extraction. This method uses two classes (yes and no) and mapping is established for every feature to one of the class.	Candidate list of features has been obtained hence extensive coverage of features are possible. Orthography is used to determine if the word is a feature or not.	Only the features that are explicitly stated by the customer are handled and the implicit features are ignored.
13	Collocation driven opinion mining method Which uses syntactic features and features related to products	Polarity for the attributes is assigned and the frequency of occurrence of the attributes is considered for analysis.	This paper does not cover any mechanism to extensively cover the entire feature set of the products.
14	Redopalfeaturebasedopiningmining	Featuresareextractedafterpos tagging and a	This classification should have

	method	feature scoring approach is used.	been done on the basis of words present in the review.
15	Feature based opinion mining	Feature mining on the basis of the most related and important feature is done by using the association rules.	Related feature may get pruned if the threshold for pruning is kept high

# **5. CONCLUSIONS**

We have elaborated some basic concepts of opinion mining used in social network in this paper. We have illustrated the basic concepts, current challenges and comprehensive study in different sections. We hope the ideas and concepts described in this paper will help us in future to study about opinion mining and sentiment analysis in social networks more deeply.

### REFERENCES

[1]. M. Jamali and H. Abolhassani, "Different Aspects of Social Network Analysis", Proceeding of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 66–72, Dec. 2006.

[2]. Chen H, Yang C, Chau M, Li S., In Intelligence and Security Informatics, Springer, Berlin, 2009.

[3]. Colbaugh, Richard; Glass, Kristin, "Estimating sentiment orientation in social media for intelligence monitoring and analysis", Proceedings of 2010 IEEE International Conference on Intelligence and Security Informatics, Vancouver, 2010.

[4]. Hatzivassiloglou, V. and K. Mckeown, Predicting the semantic orientation of adjectives, Proceedings of the 8th conference on European chapter of the Association for Computational Linguistics, pp. 174–181, 1997.

[5]. Turney, P. and M. Littman, "Measuring praise and criticism: inference of semantic orientation from association". ACM TOIS 21(4), 315–346, 2003

[6]. Wiebe, J. Learning subjective adjectives from corpora., AAAI-2000.

[7]. Harb, A., G. Dray, M. Plantié, P. Poncelet, M. Roche, and F. Trousset, "Détection d'opinion

: apprenons les bons adjectifs !", Atelier Fouille des Données d'Opinions (FODOP08), pp. 59–66, 2008.

[8]. Hu, M. and B. Liu, "Mining and summarizing customer reviews", KDD '04: Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177. ACM, 2004

[9]. Fisher, D., M. Smith, and H.Welser, "You are who you talk to: Detecting roles in usenet newsgroups", Proceedings of the 39th Annual HICSS. IEEE Computer Society, 2006

[10]. Java, A., X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities", WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD

and 1st SNA-KDD 2007 workshop on Web Mining and Social Network Analysis, pp. 56–66, 2007.

[11]. Scripps, J., P.-N. Tan, and A.-H. Esfahanian, "Node roles and community structure in networks", WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web Mining and Social Network Analysis, pp. 26–35. ACM, 2007

[12]. Maurel, S., P. Curtoni, and L. Dini, "L'analyse des sentiments dans les forums", Atelier Fouille des Données d'Opinions (FODOP 08), 2008.

[13]. D. Maynard and A. Funk, "Automatic detection of political opinions in tweets", The Semantic Web: ESWC 2011 Selected Workshop Papers, Lecture Notes in Computer Science, Springer, 2011.

[14]. D. Maynard and M. A. Greenwood, "Large Scale Semantic Annotation, Indexing and Search at The National Archives", Proceedings of LREC 2012, Turkey, 2012.

[15]. Sajin. S. Chandran, Murugappan S., "A Review on Opinion Mining from Social Media Networks", European Journal of Scientific Research, pp.430-440, 3rd October, 2012.