



Opinion Mining and Review Spam Detection: Issues and Challenges

Siddu P. Algur, Jyoti G. Biradar

Department of Computer Science, Rani Channamma University, Belagavi,
Karnataka, IndiaDOI: [10.23956/ijarcsse/V7I1/0170](https://doi.org/10.23956/ijarcsse/V7I1/0170)

Abstract— Public opinions and experience are valuable information in decision making process. Opinions may be sensitive since they may reflect one's perspective, understanding, particular feelings, way of life, and desires. Several websites encourage users to express and exchange their views, suggestions and opinions related to product, services, polices, etc. With the development of internet, people are more likely to express their views and opinions on e-Commerce sites, forums and blogs and are able to interact intensively with the internet. Product reviews composed collaboratively by many independent internet reviewers can help consumers make purchase decisions and enable enterprises to improve their business strategies. As the number of reviews is increasing exponentially, opinion mining and retrieval techniques are needed to identify important reviews and opinions to answer users queries. Most opinion mining and retrieval approaches try to extract sentimental or bipolar expressions from a large volume of reviews. It is well-known that many online reviews are not written by genuine users of products, but by spammers who write fake reviews to promote or demote some target products. Therefore, this paper presents a review of literatures that discusses about essentials of spam and review spam and brief conceptualization of opinion review mining from data mining perspective. The study finally discusses the application areas, research challenges and research scope explored from the open issues in this area.

Keywords— Customer reviews, Opinion mining, Spam, Review spam, Research Challenges.

I. INTRODUCTION

Today e-Commerce popularity has made web an excellent source of gathering customer reviews/opinions about a product that they have purchased. Thanks to the technological revolution that has accompanied the web 2.0. The web has dramatically changed the way that people express themselves and interact with others. They can now post reviews of a product at merchant sites and express their views and interact with others via blogs and forums. Such user-generated content on the web provides vital information on these products which help potential customers to find opinions of existing users before deciding to purchase a product [1]. It is now a common practice for e-Commerce web sites to enable their customers to write reviews of products that they have purchased. The reviews are then used by potential customers to find opinions of existing users before purchasing the products. They are also used by manufacturers to detect problems in their products and/or to find competitive intelligent information about their competitor [2][3]. The web contains a wealth of opinions about products, politicians and more, which are expressed in news group posts, review sites and elsewhere. As a result, the problem of 'Opinion mining' has seen increasing attention from last few years [4]. The number of customer reviews that a product receives is growing at a very fast rate. An important issue related to the trustworthiness of online opinions has been neglected most often. There is no reported study on assessing the trustworthiness of reviews, which is crucial for all opinion based applications, although web spam and email spam have been investigated extensively. As internet has no quality control, anyone can write anything on the web which results in many low quality reviews, and worse still review spam which is often biased and may mislead the customer affecting his buying decisions. Thus, it is very essential to have a mechanism which is capable of assessing the trustworthiness of reviews for proper decision making or for marketing intelligence. Trusted customer reviews are useful for both potential buyers and product manufacturers. It is more convenient and less time consuming for buyer to see at a glance feature by feature comparison of reviews written by most of the customers in taking buying decisions without getting biased and product manufacturer gets to know strengths and weaknesses of his/her own products and also that of the competitors, consumer preferences and interests by which profits could be maximized [5]. Thus, it is essential to identify trusted customer reviews.

1.1 Opinion Mining

Opinion mining is a type of natural language processing for tracking the mood of the public about a particular product. Opinion mining, which is also called sentiment analysis, includes building a system to collect and categorize opinions about a product. Automated opinion mining often uses machine learning, a type of artificial intelligence (AI), to mine text for sentiment. It has attracted a lot of researchers from different areas of research including natural language

processing, data mining, machine learning, linguistics, and even social science. Opinion mining can be useful in several ways. It can help marketers evaluate the success of an ad campaign or new product launch. For example, a review on a website might be broadly positive about a digital camera, but be specifically negative about how heavy it is. Being able to identify this kind of information in a systematic way gives the vendor a much clearer picture of public opinion than surveys or focus groups do, because the data is created by the customer [11]. In general, opinions can be expressed on anything, e.g., a product, a service, a topic, an individual, an organization, or an event. The general term object is used to denote the entity. An object has a set of components (or parts) and a set of attributes. Each component may also have its sub-components and its set of attributes, and so on. Thus, the object can be hierarchically decomposed based on the part-of relationship. [12]. An opinion has three main components, i.e., the opinion holder or source of the opinion, the object about which the opinion is expressed and the evaluation, view or appraisal, that is, the opinion. For opinion identification, all of these components are important. While opinions can be collected from different sources, like individual interactions, newspapers, television, internet etc. The Internet has become the richest source of opinion collection. Before the world wide web (www), people collected opinions manually. If an individual was to make a decision, he/she typically asked for opinions from friends and family members. To acquire public opinion, organizations often conducted surveys through focused groups. This type of survey, however, was expensive, time consuming and laborious. Now a days, the Internet provides many information and public opinions related to any product at very little cost.

A review is a sequence of sentence expressing opinions on a particular product. Various websites provides different formats for writing the reviews. There are three different types of review formats available on the web. Format 1. Pros and Cons: The reviewer is to describe Pros and Cons separately. Cnet.com uses this format. Format 2. Pros, Cons and detailed review: The reviewer is to describe Pros and Cons separately and write a detailed review. Epinions.com uses this format. Format 3. free format: The reviewer can write freely, i.e., no separation of Pros and Cons. Amazon.com uses this format [9]. The present paper discusses about the trends of Spam Analysis with respect to online opinion review mining. Section 2 discusses about the essentials of Spam Analysis and its related work. Section 3 discusses about Prior studies related to opinion mining and review spam analysis Section 4. discusses about Application areas of Online opinion mining and spam analysis Section 5. discusses about the Research Challenges in opinion mining.. Section 6 gives the Conclusion and future work.

II. ESSENTIALS OF SPAM ANALYSIS

• Spam

Unsolicited, anonymous, commercial and mass email messages, called spam are viewed as a serious problem for Internet, content quality and trust [6]. Spam is flooding the Internet with many copies of the same message, in an attempt to force the message on people who would not otherwise choose to receive it. Most spam is commercial advertising, often for dubious products, get-rich-quick schemes or quasi-legal services and customer review spam. Spamming is any deliberate action solely in order to boost a web page's position in search engine results, incommensurate with page's real value [7]. The term spamming (also, spam indexing) is used to refer to any deliberate human action that is meant to trigger an unjustifiably favourable relevance or importance for some web page, considering the page's true value. The adjective 'spam' is used to mark all those web objects (page content items or links) that are the result of some form of spamming. People who perform spamming are called spammers [8].

• Email Spam

Email Spam is any email that was not requested by a user but was sent to that user and many others, typically (but not always) with malicious intent. The source and identity of the sender is anonymous and there is no option to cease receiving future emails.

• Web Spam

Web Spam is the web pages that are the result of spamming. Web spam is the deliberate manipulation of search engine indexes. It is one of the search engine optimization methods. Implementing web spam on a search engine reduces the redundant and non-desirable results [7,8] .

• SMS Spam

SMS spam is any unwanted text message received on a mobile device. Like mail spam, SMS spam can range from unsolicited advertising to social engineering hoaxes to harmful attempts to steal subscribers personal and financial details.

• Review Spam

A review spam is considered to be totally unrelated, untrustworthy & untruthful user opinions on products and services. Such spam reviews are widespread on merchant's site and can be very harmful. For e.g a spam review that praises a product that every reviewer likes (gives a high rating) is not very damaging. However a spammer can carefully craft a review that criticizes a product that most people like can be very harmful affecting the customers buying decisions. Detecting untruthful opinion spam by manual reading is very hard, if not impossible, because a spammer can carefully craft a spam review to promote a target product or to damage the reputation of another product that is just like any other innocent review. Also the occurrence of the large number of duplicate and near-duplicate reviews written by the same reviewers on different products or by the different reviewers on the same products or different products are almost certainly considered as untruthful opinion spam reviews [1] .

• Opinion Spamming

Opinion spamming refers to writing fake reviews that try to deliberately mislead human readers or automated opinion mining systems by giving undeserving positive opinions or unjust or false negative opinions to promote or

demote some target products [9]. The opinion information also benefits the business organizations. They can monitor the consumers expressions, and effectively adjust their production and marketing strategies [10].

III. PRIOR STUDIES

The important part to gather information always seems as what the people think. The growing availability of opinion rich resources like online review sites and blogs arises as people can easily seek out and understand the opinions of others. Users express their views and opinions regarding products and services. These opinions are subjective information which represents user's sentiments, feelings or appraisal related to the same [21]. Web is widely used to interact with each other and to express different things as posting reviews of product, movies, etc., that means web users giving their own opinions. These opinions or reviews have importance in both users and vendors application. There is a need to develop a user application that gives exact result, i.e. opinion about a product. It becomes helpful to the customer or user to take a decision to buy the product. Furthermore, on vendor side, opinions are important for management of reputation and brand perception of product. Reputation management means finding out overall view about existing brand. Brand perception means how brand perceived by the customer. Opinions are beneficial to both user and vendors if the opinion posted are appropriately without any wrong intention [22]. Today, more and more people rely on the wealth of information available on the World Wide Web, and thus, increased exposure on the web may yield significant financial gains for organizations. Often, search engines are the entry ways to the web. That is why some people try to mislead search engines, so that their pages rank high in search results, and thus, capture user attention. Hence, just as with emails, one can find ways about attempts of spamming the content of the web. The first step for people to express their views online without any intervention is to find the way of detecting spam through online review. Major issues with e-commerce(online working)include spammers and different types of spam like web spam ,email spam etc. Product reviews exist in a variety of forms on the web sites dedicated to a specific type of product (such as MP3 player or movie pages), sites for newspapers and magazines that may feature reviews (like Rolling Stone or Consumer Reports), sites that couple reviews with commerce (like Amazon), and sites that specialize in collecting professional or user reviews in a variety of areas (like C_net or ZDnet in electronics, or the more broad epinions.com and Rateitall.com) [23]. The Web has dramatically changed the way people express themselves and interact with others. They can now post reviews of products at merchant sites and express their views and interact with others via blogs and forums. Online shoppers use customer's reviews to make informed buying decisions [24]. Consumer actively seeks out and reads customer reviews prior to making a purchase decision. Those studies are able to identify certain types of spammers, e.g., those who post many similar reviews about one target entity. However, in reality, there are other kinds of spammers who can manipulate their behaviors to act just like genuine reviewers, and thus cannot be detected by the available techniques [25]. Spam became one of the most significant problems for internet communication and users today. Traditional filtering methods and signature-based spam blocking systems that have come into prominence recently fall short as statistics demonstrate that spam further mount up day after day [26]. Merchants selling products on the web often ask their customers to review the products that they have purchased and the associated services. As e-commerce is becoming more and more popular, the number of customer reviews that a product receives grows rapidly. For a popular product, the number of reviews can be in hundreds or even thousands. This makes it difficult for a potential customer to read them to make an informed decision on whether to purchase the product. It also makes it difficult for the manufacturer of the product to keep track and to manage customer opinions. For the manufacturer, there are additional difficulties because many merchant sites may sell the same product and the manufacturer normally produces many kinds of products [27]. As numerous on-line product reviews that vary in quality are published every day, much attention is being paid to quality assessment of such reviews. On-line product reviews have become one of the most powerful guidelines for customer's decision making in the sense that they contain opinions or even emotions about the products directly reflecting the experiences of the old customers [28]. Product reviews are an increasingly important type of user generated online content since they offer valuable information that helps product designers better, understand the needs and preferences of consumers, and in the meantime, influences potential consumers in their purchase decision making [29]. Starting from 2003–2004, there had emerged a stream of research efforts on analyzing such online reviewers [30]. It is now known as opinion mining and the major research focuses are product feature identification and sentiment analysis. These reviews are written entirely based on the willingness of consumers, from their angle of interests, in their language and without any prescribed questions to lead them. In the past few years, there has been an obviously rising interest in the topic of automatic parsing and analysis of online reviews in several major research forums. Hu and Liu proposed a method that uses various word features, including occurrence frequency, part-of-speech tags and synonyms set in WordNet [31]. While they called it a summarization, their basic idea was to identify a noun word and its nearest opinion word. Ding et al. proposed a holistic lexicon based approach [32]. They utilized external evidence and linguistic conventions to identify the semantic orientation of opinions, and later in a further work on product entity discovery and entity assignment. Recently, a lexicalized hidden Markov model-based learning framework [33] has also been reported. Other related studies include sentiment analysis and subjective classification, for example, using a word sentiment classification approach [34], a latent semantic analysis-based approach where the cosine distance is introduced and some syntactic information in the feature sets of a support vector machine [35]. Online store reviews are an important resource to help people make wise choices for their purchases. Due to this reason, the review system has become a target of spammers who are usually hired or enticed by companies to write fake reviews to promote their products and services, and/or to distract customers from their competitors [36]. Driven by profits, there are more and more spam reviews in major review websites, such as PriceGrabber.com, Shopzilla.com, or Resellerratings.com. Spammers are starting to corrupt the online review system and confuse the consumers [37]. A thorough study of supervised learning approaches for deceptive review detection was

conducted in [38]. They studied how well existing research methods work for detecting real-world fake reviews on a commercial website. The authors tested their models using the Amazon Mechanical Turk (AMT) synthetic fake reviews dataset on a real-world fake reviews dataset procured from Yelp. In this study, they found similar results to previous studies, confirming that using n-gram features performs well on the AMT dataset, however, when used with the real world Yelp dataset it performed significantly worse. They observed that using behavioral features yields higher performance than linguistic features alone on the real world Yelp dataset. Three different features sets were used in the experiment: LIWC, POS and bigrams. In addition, feature selection using Information Gain (IG) was applied to select the top 1 and 2 % features. One of the main conclusions of the study was that the synthetic reviews are not necessarily representative of what is found in real world review spam. All of these contributions investigate ways to classify a review as positive, negative, or neutral at different levels or from particular perspectives, namely, the word level, the sentence level, the document level and the feature level. A novel technique for detecting review spammers was proposed where they exploit the “bursty” nature of reviews generated by spammers to identify review spam [39]. Bursty reviews are reviews that suddenly become popular and receive great attention from reviewers within a certain time period or certain area. The reviews and reviewers in those situations become suspicious as review spam and review spammer respectively. For burst detection, the authors used Kernel Density Estimation (KDE) techniques to detect review bursts. KDE is a technique closely related to histograms, which has attributes that allow it to asymptotically converge to any density function. Behavioral features for spammers were created that combined the spammers’ behaviors with the features of review bursts. In addition, these features can be used in conjunction with review spam features in a hybrid approach to improve the classification results. The experiment conducted in [40] is a notable exception, as they used Information Gain (IG) to perform feature selection of top 1 and 2 % of features. Though they found this had no impact on classifier performance, one believed that using feature selection techniques one can potentially improve performance based on results from other domains. Feature selection also has the benefit of reducing the computational costs associated with training a classifier. This is highly desirable as review spam detection is a big data domain and datasets may have a very large number of instances and features. Finally, there are a massive number of online reviews, and fake reviews are usually less frequent than truthful ones, resulting in highly imbalanced datasets. Class imbalance can adversely affect classifier performance as the majority class may be favoured, and must be taken into consideration when training a model. Two works have considered the class imbalance problem in this domain, [41] and [42]. Both used random under sampling and random oversampling to overcome imbalanced distributions and have promising but inconclusive results. Ensemble techniques can be used alongside, or in place of, data sampling as they have been shown to be more robust to the effects of class imbalance than single classifiers. The experiments conducted in [49] used frequent item set mining method in spam detection in the collaborative setting to discover fake reviewer groups. It was first method to use a frequent item set mining method to find a set of candidate groups and then used several behavioral models derived from the collusion phenomenon among fake reviewers and relation models based on the relationships among groups, individual reviewers, and products they reviewed to detect fake reviewer groups. To assess proposed method, authors built a labeled dataset using expert human judges which is taken from Amazon dataset and built a labeled dataset of fake reviewer groups. In [50], a novel review spam detection method was used, which is underpinned by an unsupervised inferential language modeling framework. In this work, each review is represented by a TFIDF vector, and then every pair of reviews of a product category was based on the cosine similarity measure. A support vector machine was also applied to classify untruthful review and all default parameters of the SVM light package were used.

IV. APPLICATION AREAS OF OPINION MINING AND REVIEW SPAM DETECTION

Since the opinion based or feedback based applications are more fashionable, now a days, the natural language processing community shows much interest in Spam Analysis and Opinion Mining system. The explosion of internet has changed the people’s life style, now they are more expressive on their views and opinions [13] [14], and this tendency helped the researchers in getting user-generated content easily. The major applications of online opinion review mining and spam analysis are the following:

1) Purchasing Product or Service : While purchasing a product or service, taking right decision is no longer a difficult task. By online reviews, people can easily evaluate other’s opinion and experience about any product or service and also he can easily compare the competing brands. Now people don’t want to rely on external consultant. The opinion mining analysis extracts people’s opinion from the huge collection of unstructured content, the internet, and analyzes it and then present to them in highly structured and understandable manner.

2) Opinion spam detection : Since internet is available to all, anyone can put anything on internet, which increases the possibility of spam content on the web. People may write spam content to mislead the consumer. Opinion mining and sentiment analysis can classify the internet content into ‘spam’ content and ‘not spam’ content [14].

3) Quality Improvement in product or services : By Opinion mining and sentiment analysis the manufactures can collect the critic’s opinion as well as the favorable opinion about their product or service and thereby they can improve the quality of their product or service. They can make use of online product reviews from websites such as Amazon and C|Net [15,16], RottenTomatoes.com [17] and IMDb [18].

4) Marketing research : The on-line word-of-mouth behavior represents new and valuable sources of applications used in companies, organizations and individuals to gain more information effectively and easily.

5) Detection of flame : The monitoring of newsgroup and forums, blogs and social media is easily possible by sentiment analysis. Online opinion review mining can automatically detect arrogant words [19], over heated words or hatred language used in emails or forum entries or tweets on various internet sources.

6) Recommendation Systems: By classifying the people's opinion into positive and negative, the system can say which one should get recommended and which one should not get recommended [20].

7) Policy Making: Through Spam analysis, policy makers can take citizen's point of view towards some policy and they can utilize this information in creating new citizen friendly policy.

8) Decision Making: People's opinion and experience are very useful element in decision making process. Online opinion review mining gives analyzed people's opinion that can be effectively used for decision making.

V. RESEARCH CHALLENGES OF OPINION MINING AND REVIEW SPAM DETECTION

Large amounts of online reviews, the valuable voice of the customer, benefit consumers and product designers. Identifying and analyzing helpful reviews efficiently and accurately to satisfy both current and potential customers needs have become a critical challenge for market-driven product design. Few research challenges of opinion review mining and spam analysis are :

1) Detection of spam and fake reviews: The web contains both authentic and spam contents. For effective sentiment classification, this spam content should be eliminated before processing. This can be done by identifying duplicates by detecting outliers and by considering reputation of reviewer, but it is time consuming. Hence, new machine learning tools and algorithms have to be incorporated to detect review spam [1].

2) Limitation of classification filtering: There is a limitation in classification filtering while determining most popular thought or concept. For better sentiment classification result, this limitation should be condensed by using few collaborative filtering technique.

3) Asymmetry in availability of opinion mining software: The opinion mining software is very expensive and currently affordable only to big organizations and government. It is beyond the common citizen's expectation. This should be available to all people, so that everyone gets benefit from it.

4) Incorporation of opinion with implicit and behavior data: For successful analysis of sentiment and spam, the opinion words should integrate with implicit data. The implicit data determine the actual behavior of sentiment words. Few examples of opinion with implicit and behavior data.

- The word that is considered to be positive in one situation may be considered negative in another situation. Example the word "long" for instance. If a customer said a laptop's battery life was long, that would be a positive opinion. If the customer said that the laptop's start-up time was long, however, that would be is a negative opinion. These differences mean that an opinion system trained to gather opinions on one type of product or product feature may not perform very well on another.
- People don't always express opinions the same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In opinion mining, however, "the movie was great" is very different from "the movie was not great".
- People can be contradictory in their statements. Most reviews will have both positive and negative comments, which is somewhat manageable by analyzing sentences one at a time. However, the more informal the medium (twitter tweets or blog posts for example), the more likely people are to combine different opinions in the same sentence. For example: "the movie bombed even though the lead actor rocked it" is easy for a human to understand, but more difficult for a computer to parse. Sometimes even other people have difficulty understanding what someone thought based on a short piece of text because it lacks context. For example, "That movie was as good as his last one" is entirely dependent on what the person expressing the opinion thought of the previous film.

5) Domain-independence: The biggest challenge faced by opinion mining and sentiment analysis is the domain dependent nature of sentiment words. One features set may give very good performance in one domain, at the same time it perform very poor in some other domain.

6) Natural language processing overheads: The natural language overhead like ambiguity, co-reference, Implicitness, inference etc. creates hindrance in spam analysis.

7) Aspect based opinion mining : There are various challenges that makes the problem of opinion mining hard. A few of them is mentioned below:

- Identifying aspects : The first challenge in identifying aspects is that different reviewers may use different words or phrases to express the same aspect.
- Noisy information: Full text reviews normally include a large amount of irrelevant information, e.g., opinion about the manufacturer of the product and information about the reviewer.
- Identifying opinions in comparative sentences : A comparative opinion expresses a relation between two or more items and/or a preference of the reviewer based on some shared aspects of the item.
- Co-reference resolution : The Co-reference occurs when multiple expressions in a sentence or document refer to the same thing.

8) Opinions Change with Time : Another challenge lies in the issue of being able to monitor opinions changing with the passage of time. This helps us to observe if a certain product gets improved with time, or people change their opinion about a product and get convinced for it with time. To identify how the people's mood changes over time in described in the work [43].The work done observes blogs where the mood is explicitly specified either by selecting from a predefined list of moods or by entering it as free text.

9) Strength of Opinions : Identification of the strength of an opinion is another challenge faced in opinion mining. The strength of an opinion can change as the discussion continues in a forum i.e. arguments used during discussion are strong enough to change the strength. SentiWordNet application has been used to identify strength of opinions [44].

10) Sentences with Mixed Views: A bigger challenge for opinion mining comes when people express positive and negative review in the same sentence. This is mostly the issue when people are communicating through informal mediums like blogs and forums. People are more likely to combine different opinions in the same sentences. Such sentences can be difficult to parse for opinion mining. Sentiment mining or opinion mining is contrasted generally with the traditional fact-based text mining. Text mining seeks to classify documents by topics while opinion mining generalizes text across many domains and users. Strength of a feeling, degree of positivity and similar factors can be of potential importance in opinion mining [46].

11) Different People Different Writing Style: The fact that comments or views entered by people who are different from each other in the way they write, their use of language, abbreviations and their knowledge is a challenge on its own. People also do not express opinion in the same way. One might use certain negative terms in a sentence Text that appears in an online newspaper and that which appears in an online forum is widely different. The mining of online forums and discussions is a challenge on its own. Some possible reasons include the use of abbreviations, the entry of comments by different people, who differ in the way they write or in the knowledge of the language they use [15] Reference [25] addresses this problem. In this reference some recent works have started to classify product features but they heavily rely on linguistic and natural language processing techniques. However, writing in consumer reviews is usually not formal and does not follow grammatical rules which make language processing approach inappropriate to use. Therefore, the linguistic and language processing approach is not satisfactory. So, a sentiment analysis system is suggested for classifying products features in consumer reviews by means of mining class association rules.

12) Misleading Opinions due to sarcastic and ironic statements ; Sarcastic and ironic sentences exist in text. In such a scenario, positive words can have negative sense of usage in a metaphorical manner. Text in a statement can be hard to identify as sarcastic or ironic which can lead to erroneous orientation and misleading opinion mining [47].

13) Misleading Opinions due to spam opinion: Dishonest opinions or reviews intend to affect opinion mining about a product or service. Detecting such opinions is important for practical utilization of opinion mining. Semantic coverage may be useful feature for detecting spam. Review spam exists due to repeating of important terms and dumping of many unrelated terms [48].

VI. CONCLUSION & FUTURE WORK

Opinions are very important for anyone who is going to make a decision. Web mining has emerged in recent years as an attractive technology to individuals and corporations to know others opinions. Opinion mining is helpful for individuals when they want to buy a product and they can decide which product to buy, by studying the summarized opinions instead of studying long reviews and making summary themselves. Review text is an important source of information for the consumers before purchasing any product from e-Commerce websites. In recent years, review spam has received significant attention in both business and academia. This survey covers prior studies, application areas, research challenges and research scope in opinion mining and review spam analysis. Although various methods have been proposed to detect review spam, still there is scope to improve & develop sentiment word identification, to apply sentiment analysis on short sentence like abbreviations, to develop fully automatic analyzing tool, to successfully handle bi polar sentiments, effective analysis of policy / procedure opinionated content. To implement Big data and possible formulation of new knowledge discovery process and tools for analyzing and detecting review spam gives scope for future work.

REFERENCES

- [1] Arjun Mukherjee , Bing Liu , Junhui Wang, Natalie Glance, Nitin Jindal, “ Detecting Group Review Spam “ in poster March 28 –April 1,2011,Hyderabad, India.
- [2] Nitin Jindal and Bing Liu “ Analyzing and Detecting Review Spam” in ICDM 2007 IEEE.
- [3] Hu and B. Liu. 2004. Mining and Summarizing Customer Reviews. In *KDD*, pages 168–177, Seattle, WA
- [4] Ana-Maria Popescu and Oren Etzioni “Extracting Product Features and Opinions from Reviews “.
- [5] Siddu P. Algur, Amit P.Patil, P.S Hiremath, S. Shivashankar “Conceptual level Similarity Measure based Review Spam Detection” 2010 IEEE.
- [6] Pedram Hayati, Vidyasagar Potdar Toward Spam 2.0: An Evaluation of Web 2.0 Anti-Spam Methods.
- [7] Sumit Sahu, Bharti Dongre, Rajesh Vadhvani, “ Web Spam Detection Using Different Features “ in IJSCE, ISSN : 2231-2307 , Volume-1, Issue -3,July 2011.
- [8] Z.Gyongyi & H. Garcia-Molina. “Web Spam Taxonomy”. Technical Report, Stanford University, 2004.
- [9] Siddu P.Algur ,Amit P.Patil,P.S Hiremath,S.Shivashankar “ Spam Detection of Customer Reviews from Web Pages “ in IT and Business Intelligence.
- [10] Fangtao Li, Minlie Huang,Yi Yang and Xiaoyan Zhu, “ Learning to Identify Review Spam “ in IDRC 104519-006 , the Chinese Natural Science Foundation Grants No. 60973104 and 60803075.
- [11] Mukherjee A, Liu B, Glance N (2012) Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st international conference on World Wide Web. (pp. 191–200). ACM, Lyon, France
- [12] Ott M, Cardie C, Hancock JT (2013) Negative Deceptive Opinion Spam. In: HLT-NAACL., pp 497–501

- [13] Haseena Rahmath P “Opinion Mining and Sentiment Analysis -Challenges and Applications” in (IJAIEM) Volume 3, Issue 5, May 2014 ISSN 2319 – 4847.
- [14] Liu, B. (2010), “Sentiment Analysis and Subjectivity”. Appeared in Handbook of Natural Language Processing, Indurkha, N. & Damerau, F.J. [Eds.].
- [15] Dave K., Lawrence, S. & Pennock, D.M. (2003), “Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews”. In Proceedings of the 12th International Conference on World Wide Web, p. 519-528.
- [16] Hu, M. & Lui, B. (2004), “Mining and Summarizing Customer Reviews”. In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2004 (KDD-2004), p. 168–177.
- [17] Pang, B. & Lee, L. (2004), “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts”. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, p.271-278.
- [18] Pang, B., Lee, L. & Vaithyanathan, S. (2002), “Thumbs Up? Sentiment Classification Using Machine Learning Techniques”. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2002), p. 79-86.
- [19] Boiy, E., Hens, P., Deschacht, K. & Moens, M.-F. (2007), “Automatic Sentiment Analysis in OnLine Text”. In Proceedings of the Conference on Electronic Publishing (ELPUB-2007), p. 349-360.
- [20] Segaran, T. (2007), Programming Collective Intelligence. Sebastopol: O’Reilly Media, Inc.
- [21] Nidhi Mishra and C K Jha. Article: Classification of Opinion Mining Techniques. International Journal of Computer Applications 56 (13):1-6, October 2012.
- [22] Manali S.Patil, A.M.Bagade “Online Review Spam Detection using Language Model and Feature Selection “International Journal of Computer Applications (0975 – 8887) Volume 59– No.7, December 2012.
- [23] Kushal Dave , Steve Lawrence, David M. Pennock “Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews” Budapest, Hungary. ACM 1581136803/03/0005.
- [24] Minqing Hu and Bing Liu “Opinion Extraction and Summarization on the Web”.
- [25] Guan Wang, Sihang Xie, Bing Liu, Philip S.Yu, “Review Graph based Online Store Review Spammer Detection” in part by (NSF) National Science Foundation through grants IIS-0905215, OISE-0968341,DBI-0960443 and IIS-1111092.
- [26] Assist.Prof.Dr. Ahmet BARAN “Stopping spams with sending session verification “Erzincan University, Vocational High School,24100, Erzincan, Turkey.
- [27] Minqing Hu and Bing Liu “Mining and Summarizing Customer Reviews “ACM 1-58113-888-1/04/0008.
- [28] Hye-Jin Min, Jong C. Park “Identifying helpful reviews based on customer’s mentions about experiences “Expert Systems with Applications 39 (2012) 11830–11838, Elsevier
- [29] Ying Liu ,Jian Jin, Ping Ji, Jenny A. Harding, Richard Y.K. “Identifying helpful online reviews: A product designer’s perspective “Computer-Aided Design 45 (2013) 180–1940010-4485/\$ –2012 Elsevier .
- [30] Liu Y, Lu WF, Loh HT. Knowledge discovery and management for product design through text mining—a case study of online information integration for designers. In: Proceedings of the 16th international conference on engineering design. ICED’07. Paris, France. 2007. 329 (1–12).
- [31] Ding X, Liu B. The utility of linguistic rules in opinion mining. In: Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval. Amsterdam (The Netherlands): ACM; 2007.
- [32] Hu M, Liu B. Mining opinion features in customer reviews. In: Proceedings of the 19th national conference on artificial intelligence. San Jose, USA. 2004.p. 755–60.
- [33] Ding X, Liu B, Zhang L. Entity discovery and assignment for opinion mining applications. Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. Paris (France): ACM;2009
- [34] Jin W, Ho HH. A novel lexicalized HMM-based learning framework for web opinion mining. In: Proceedings of the 26th annual international conference on machine learning. Montreal (Quebec, Canada): ACM; 2009.
- [35] Kim S-M, Hovy E. Determining the sentiment of opinions. In: Proceedings of the 20th international conference on computational linguistics. Geneva, Switzerland: Association for Computational Linguistics. 2004.
- [36] Turney PD, Littman ML. Measuring praise and criticism: inference of semantic orientation from association. ACM Trans Inf Syst (TOIS) 2003;21:315–46.
- [37] Mullen T, Collier N. Sentiment analysis using support vector machines with diverse information sources. In: Proceedings of conference on empirical methods in natural language processing. EMNLP. 2004. p. 412–8.
- [38] Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn pp:1157–1182
- [39] Fei G, Mukherjee A, Liu B, Hsu M, Castellanos M, Ghosh R (2013) Exploiting Burstiness in reviews for review spammer detection. ICWSM 13 pp: 175–184.
- [40] Mukherjee A, Venkataraman V, Liu B, Glance NS (2013) What yelp fake review filter might be doing? Boston, In ICWSM.
- [41] Hammad ASA An Approach for Detecting Spam in Arabic Opinion Reviews. Doctoral dissertation, Islamic University of Gaza in 2013.
- [42] Al Najada H, Zhu X (2014) iSRD: Spam review detection with imbalanced data distributions. In: Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on (pp. 553–560). IEEE, San Francisco,

- [43] Krisztian Balog, Maarten de Rijke Decomposing Bloggers' Moods - Towards a Time Series Analysis of Moods in the Blogosphere 2006.
- [44] B. Ohana, B. Tierney Opinion Mining with SentiWordNet .2011.
- [45] Andreas Auinger, Martin Fischer. Mining consumers' opinions on the web. 2008.
- [46] Hyun Duk Kim, ChengXiang Zhai.2009.Generating comparative summaries of contradictory opinions in text.
- [47] Zhongwu Zhai, Bing Liu. Identifying Evaluative Sentences in Online Discussions. 2011.
- [48] Bo Pang and Lillian Lee. Opinion Mining and Sentiment Analysis, Department of Computer Science University of Illinois at Chicago 2008.
- [49] A. Mukherjee, B. Liu, N. Glance, —Spotting Fake Reviewer Groups in Consumer Reviewsl, WWW 2012, April 16–20, 2012, Lyon, France , ACM
- [50] C.L. Lai, K.Q. Xu, Raymond Y.K. Lau and Yuefeng Li, —High Order Concept Associations Mining and Inferential Language Modeling for Online Review Spam Detectionl, 978-0-7695-42577/10 , IEEE, 2010 .