International Journal of Information & Computation Technology.
ISSN 0974-2239 Volume 4, Number 16 (2014), pp. 1625-1635

http://www. irphouse.com

Opinion Mining: A Review

Gautami Tripathi[1] and Naganna S[2]

*[1]PG Scholar, School of Computing Science and Engineering,
Galgotias University, Greater Noida, Uttarpradesh.
[2]Assistant Professor, School of Computing Science and Engineering,
Galgotias University, Greater Noida, Uttarpradesh.
[1]gautami1489@gmail.com [2]nsc.chetty@gmail.com*

Abstract

The rapid development of the web applications has created numerous opportunities for the people to freely voice their opinions. The explosion of social media and other web applications such as micro-blogging, e-commerce sites, news portals etc. has resulted in large quantities of user generated content in the form of comments, reviews, feedback, recommendations and ratings. How to analyze and summarize the views expressed in such large opinioned text is a new and fast growing field for research. Sentiment analysis of the user generated content can be very useful for the business organizations, individuals and the Government. In the past years sentiment analysis and opinion mining has emerged as one of the popular techniques for information retrieval and web data analysis. This paper presents a survey on sentiment analysis and the related techniques. It also discusses the application areas and challenges for sentiment analysis with insight into the past researches.

Keywords: Opinion mining, Sentiment analysis and opinions.

1. Introduction

Opinions of others highly influence the human behavior and are central to almost all decision making activities. The major part of our information gathering process is to find out what others think. While making any online transaction the customer usually checks the comments and reviews posted by the other customers. The fast growing social web has significantly contributed to this user generated data including reviews, comments, opinions, services and events. This data is useful for the customer as well as the manufacturer. The manufacturers can get a reality check about their product strength and weaknesses based on the sentiments of the customer. How to analyze and summarize the views expressed in such large opinioned text is a new growing field for research. This new research domain is called opinion mining. The area of opinion mining, also called sentiment analysis is concerned with analyzing people's opinions, sentiments, emotions, evaluations and attitudes towards objects such as events, organizations, products, issues and their features.

2. Evolution

Natural language processing (NLP) and computational linguistics have a rich historical background but still there was minimum research in the field of opinion mining before the year 2000 except for some earlier work on subjectivity, interpretation of metaphors, sentiment adjectives and viewpoints [14, 15, 33, 34]. Apart from its applications in the field of data mining, web mining, and information retrieval, the sentiment analysis task has spread to the management sciences. Insights and applications from sentiment analysis have been useful in other areas including politics, law making, sociology and psychology.

The term sentiment analysis and opinion mining were first introduced in 2003 by Nasukawa et al and Dave et al respectively [22, 9]. However the researches in the field of sentiments and opinions started earlier. J.M. Wiebe [33] presents an algorithm to identify the subjective characters in fictional narrative text based on the regularities in the text. Later J.M. Wiebe [34] performed extensive examination to study the naturally occurring narratives and regularities in the writings of authors and presents an algorithm that tracks the point of view on the basis of these regularities. M.A. Hearst [15] defines a direction based text interpretation approach for text based intelligent systems to refine the information access task. Hatzivassiloglou and McKeown [14] proposed a method to find the semantic orientation of the adjectives and predicted whether two conjoined adjectives are of same polarity with 82% accuracy. L.Terveen et al [18] designed an experimental system, PHOAKS (people helping one another know stuff), to help users locate information on the web. The system uses a collaborative filtering approach to recognize and reuse recommendations.

The research in the field of opinion mining has grown rapidly after the year 2000. The major reason behind this explosive growth is the World Wide Web. Earlier there was little opinionated text for study but today there is large amount of user generated content in the form of comments, reviews and debates. J.M. Wiebe [35] uses the results of a clustering method (Lin, 1998) to identify strong clues of subjectivity in corpora. Initially the adjective features were developed using the clustering method which were then refined by adding the lexical semantic features of adjectives. J.Tatemura [17] developed a browsing method using virtual reviewers for the collaborative exploration of movie reviews from various viewpoints. S. Morinaga et al [31] worked in the area of marketing and customer relationship management and presented a framework for mining product reputation on internet. The defined approach automatically collects the user's opinions from the web and applies text mining techniques to obtain the reputation of the products. P.D. Turney [25] presents an unsupervised method to classify the reviews as thumbs up (recommended) or thumbs down (not recommended). It uses document level sentiment classification and

Pointwise Mutual Information to obtain the average semantic orientation of the reviews. The algorithm achieves an average accuracy of 74% for 410 reviews. Later Turney and Littman [28] expanded the work by presenting an approach to find out the semantic orientation of a text by calculating its statistical association with a set of positive and negative words using Pointwise Mutual Information (PMI) and Latent Semantic Analysis (LSA). The method when tested with 3596 words (1614 positive and 1984 negative) achieves an accuracy of 82.8%. B. Pang et al [4] perform document level sentiment classification using standard machine learning techniques. They used Naïve Bayes, Maximum Entropy and SVM techniques to obtain the results for unigrams and bigrams. Dave et al [9] trained a classifier using reviews from major websites. The results obtained showed that by changing a few steps in the process higher order n-grams can give better results than unigrams. Das and Chen [10] define an algorithm that consists of different classifier algorithms coupled together for extracting sentiments from stock message boards. Hu and Liu [19] worked on feature level opinion mining and determined the polarity of the object features without considering the strength of the opinions. Their work focused on mining and summarizing the customer reviews. Hu et al [16] derives an analytical model to examine whether the online review data reveals the true quality of the product. Liu et al [39] presented a model to predict the sales performance. They proposed S-PLSA for predicting the sentiments in the blogs and using S-PLSA as a means to summarize the sentiments developed ARSA (auto regressive sentiment aware) model to predict the sales performance based on the sentiments and product's past performances. Dellaroca et al [11] uses the case of Motion Pictures to explore the value of online product reviews in forecasting sales. They proposed diffusion models to capture the unique aspects of the entertainment industry and test their performance in context of very early post -release revenue forecasting. Ghose et al [13] uses econometrics to identify the economic value of text showing that user feedback affects the pricing power of merchants which can be measured to determine the polarity and strength of the user feedback. Park et al [26] examines the involvement of the online customer reviews in affecting the purchasing intentions. Archak et al [2] uses product demand as the objective function and derive a context aware interpretation of opinions showing how the opinions affects the user's choice. Chen and Xie [40] define online customer views as a new element in the marketing communication mix. They studied the role of customer reviews in marketing. Ding et al [37] proposed a holistic approach to infer the semantic orientation of an opinion word based on review context and combine multiple opinions words in same sentence. A system named Opinion Observer was also implemented based on the proposed technique. Murthy G. and Bing Liu [21] proposed a method which study sentiments in comparative sentences and also deals with context based sentiments by exploiting external information available on the web. W.Y. Kim et al [36] proposed a method for opinion mining using association rules. They proceed by POS tagging of the review sentences followed by the extraction of the feature and opinion words in the form of transaction. The transaction data was analyzed using the association rule mining methods and then use PMI to summarize the discovered association rules. Carvalho et al [27] presented a set of clues to detect irony in the positive sentiments. They were successful in identifying the ironic sentences with a relatively high precision ranging from 45% to 85%. Ding and Liu [38] used the supervised learning approach to present the problem of object and attribute co-reference. Paul et al [20] suggested a comparative LexRank approach to summarize contrastive viewpoints in opinionated text. Tumasjan et al [32] used twitter sentiments to predict election results. They analyzed over 100000 twitter messages using LIWC text analysis software and concluded that twitter can be seen as a valid real time indicator of political sentiments. Davidov et al [12] used a semi-supervised sarcasm identification algorithm (SASI) to identify sarcasm. Zhongwu Zhai et al [42] modeled a semi-supervised learning method to study the problem of product feature clustering for opinion mining and proposed to use the EM algorithm to solve it. Zhai et al [41] also proposed an unsupervised approach to identify the evaluative sentences in the online discussions. Bollen et al [8] used the twitter moods to predict the stock markets. A. Kamal et al [1] implemented a rule-based system to mine product features, opinions and their reliability scores. The proposed system uses linguistic and semantic analysis of text to mine the feature-opinion pairs from review documents. The graph based ranking algorithm HITS (Hyperlink-Induced Topic Search) was then used to obtain the reliability scores for the feature-opinion pair and the corpus.

3. Opinion Mining

Textual data can be classified into two broad categories namely- facts and opinions [7]. Unlike the factual information that is represented by objective statements, opinions and sentiments are characterized by the subjective nature. Opinions can be of different types- *direct*, *indirect* and *comparative* [5]. Direct opinion expresses the direct views about an object. For example, *"The drawing is beautiful."* Indirect opinion indirectly expresses the views about the object. For example, *"I had a headache after watching the movie"* gives a negative review about the movie. Comparative opinions describe the likes and dislikes of the opinion holder about an object over the other. For example, *"Samsung mobiles are better than Nokia."*

An opinion consists of two key components, *object* or *object feature (o)* and *sentiment* or *opinion* about the object(s) i.e. *(o, s)* where *o* is an entity or aspect of entity about which the opinion has been expressed and *s* is the positive, negative or neutral sentiment about *o*. Another factor defining the strength of the opinion (high, medium or low) can also be associated with it. This definition, although concise may not be appropriate for defining the online reviews of products as the complete details of the target product may be complicated. We use the following review to enlighten the problem. The sentences are numbered for easy reference.

Posted by: John *dated: April 18, 2014*

(1) I bought a Samsung GalaxyS3 a week ago.
(2)I like it.
(3)The display is amazing.
(4)The battery life and camera quality is also good.
(5)However, I think it's overpriced."

This review presents a number of opinions about Samsung GalaxyS3. Sentence (2) expresses a positive opinion about GalaxyS3 as a whole. Sentence (3) expresses a positive opinion about its display. Sentence (4) expresses a positive opinion about its battery life and camera quality. Sentence (5) expresses a negative opinion about its pricing. Here in sentence (3) the opinion target is actually the display of Samsung GalaxyS3. Thus, we can define an entity as a pair *(T, W)* where *T* represents the parts or sub-parts and *W* represents the set of attributes. For the above review GalaxyS3 is an entity. Its parts include battery, camera, display screen and so on. Each part has its own set of attributes like picture quality for camera, battery life and battery weight or battery and so on. Now we can redefine the opinion as a quintuple $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ [5] where e_i is the name of the entity, a_{ij} is the aspect of the e_i, s_{ijkl} is the sentiment on aspect a_{ij} of entity e_i, h_k is the opinion holder and t_l is the time when the opinion is expressed by h_k. Here e_i and a_{ij} together represents the opinion target. This definition covers almost all the possible facets of semantic meaning of an opinion. Fig.1 shows the architectural view of opinion mining. The process of opinion mining begins with collecting the online data from public forums like blogs, product review sites and discussion boards or from private logs using Facebook and Twitter. Generally the collected data is bulky and noisy so it is cleaned before the analysis is performed.

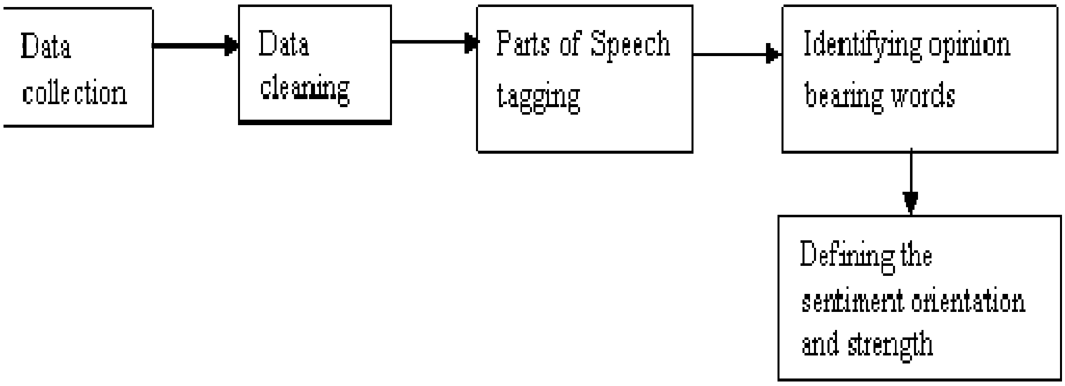

Fig.1 Architectural view of opinion mining.

The actual opinion mining begins with identifying the opinion bearing words followed by finding their semantic orientations and strength. The task of opinion mining can be performed at three different levels namely, *document level, sentence level* and *aspect level*. At document level the entire document is classified as positive, negative or neutral [25, 4]. This level of analysis assumes that the entire document expresses opinions on single entity (object). It is not applicable to documents evaluating multiple objects. At sentence level we analyze each sentence and determine its polarity (positive, negative or neutral) [29, 18]. At aspect level we perform a finer analysis by identifying and extracting the product features from the source data [19].

Presently, the market is filled with a number of tools that performs opinion mining on the available datasets. Some examples of currently available tools are:

1. Twitrratr - www.twitrratr.com
2. Sentiment 140 - http://www.sentiment140.com
3. Tweetfeel - www.tweetfeel.com
4. Opinmind - www.opinmind.com
5. Social Mention - www.socialmention.com

4. Techniques for Sentiment Classification

4.1. Sentiment classification using supervised learning: Supervised Learning is implemented by building a classifier. It requires two sets of documents, training set and test set. This technique is also called the Machine Learning based method. The classifier is trained by examples which can be manually labeled. The most frequently used algorithms for supervised sentiment classification are support vector machines (SVM), Naive Bayes classifier and Maximum entropy. Pang, Lee and Vaithyanathan [4] was the first paper to take this approach to classify movie reviews into two classes, positive and negative. One major task for the sentiment classification is to choose the appropriate set of features for classification. The most commonly used features include term presence and their frequency, parts of speech, negations and opinion words and phrases.

4.2. Sentiment classification using unsupervised learning: In the unsupervised sentiment classification the text is classified by comparing it against the word lexicons. The sentiment value of these word lexicons is determined prior to the sentiment analysis. These sentiment lexicons define the collection of words and expressions that are used to express people’s feelings, views and opinions. To better understand this we can initially begin with a positive or negative word lexicon. The document is analyzed and checked or the presence of positive and negative word lexicons. If the document has more positive word lexicons it is considered to be a positive document else if more number of negative word lexicons are present the document is considered negative. A significant breakthrough in the unsupervised sentiment classification was done by Turney [25]. He used the words “poor” and “excellent” as the seed words and calculated the orientation of words based on them. He was successful in achieving 66% accuracy for the movie review domain.

5. Application areas

5.1. E-commerce: Many websites like amazon.com provides summary of their products and allow the users to submit their views about it. Customers can easily view information about the product as a whole or about a specific feature and can share their shopping experiences with other customers. The product manufacturers also get their feedback and can adjust their marketing strategies [19].

5.2. Voice of customer (VOC): VOC describe the in-depth process of capturing a customer's expectations, preferences and aversions. It is market research technique that defines the customer’s needs and expectations. Sentiment analysis can help in analyzing the customer’s views and thus defining the credibility and reliability of the product [19].

5.3. Government: Government can assess its strengths and weaknesses by analyzing public opinions and views on different issues [13]. A number of arguments mapping software are available that helps to logically link the policy statements. For example, Debatabase, Compendium, Debategraph and Cohere.

5.4. Marketing: Sentiment analysis provides an insight into what can be done to encourage the customers to be loyal and to purchase. Automated content analysis allows to process large amounts of data and helps to identify the relevant comments and associated polarity. Thus, it can significantly contribute towards the brand reputation management [32].

5.5. Politics: The views of voters can be analyzed to predict the results of the elections [32]. Opinion mining ensures an accurate reflection of reality by listening rather than by asking. A number of voting advice applications are available in the market that helps voters to examine the political scenario and views of other voters. For example, www.smartvote.ch.

5.6. Blog analysis: Sentiment analysis can be effectively used to mine contentions in discussions and debate forums [3]. It can be applied to analyze blog posts and perform subjectivity and classification on it.

5.7. Stock Market Prediction: Opinion mining can be efficiently used to predict the events in stock market [8].

5.8. Application as a supporting technology: Opinion mining can serve an important role in enabling technology for other systems. Recommendation systems are one such example where entities can be recommended based on the positive opinion count [17].

6. Challenges for Opinion Mining

6.1. Co-reference resolution: Co-reference resolution refers to the problem of identifying what a noun phrase or pronoun refers to. For example, "We attended the lecture and went for lunch, it was awful." Here it is difficult to identify what does "it" refers to. Xiaowen D. and Bing Liu investigated the co-reference problem in opinion mining context by studying the object and attribute resolution [38].

6.2. Sarcastic sentences: Text may contain sarcastic sentences or hidden emotions. Theses emotions are hard to identify and may lead to erroneous opinion mining. For example, *"The mobile was great, its processor crashed on the second day."* In such cases the positive words may define negative emotions [12, 27, 30].

6.3. Linguistic issues: Sometimes a single sentiment word may have opposite polarity in different contexts [5]. For example, *"long"* can imply different meaning in different contexts, e.g., *"This mobile has a long battery life,"* gives a positive opinion about the battery life but it can also imply negative sentiment, e.g., *"The engine takes long time to start."* In interrogative and conditional sentences it is often observed that the sentence containing sentiment word does not express any opinion. For example, *"Can you tell me which restaurant is good for Chinese food?"* and *"If I can find a good dress I will buy it for the prom."* In another case some sentences without any sentiment word may express opinions. For example, *"The washing machine consumes a lot of electricity."*

6.4. Opinion spamming: Social media allows people to freely express their views without disclosing their identity. Some people take advantage of this anonymity and post fake reviews to promote or discredit target products. This activity is called opinion spamming. Spam reviews can be of three types namely- untruthful opinions, opinions on brands only and non opinions [23, 24].

6.5. Volatility over time: Social media especially twitter contains opinions that vary over time. These contradictions and changes need to be tracked to capture the trends over time.

7. Conclusion

This paper presents a detailed survey on various researches in the field of opinion mining with insight into its applications and challenges. In the last two decades sentiment analysis has grown as an important area to mine large quantity of data. Although many algorithms are present that provides good results for sentiment analysis but still no technique can resolve the challenges completely. Recent approaches to opinion mining aims at handling the conceptual rules that governs sentiments. The future systems needs to combine the concepts of common sense and reasoning to present a broader aspect of opinion mining that are more deeply inspired by human thoughts and psychology.

REFERENCES

[1] A Kamal, M. Abulaish and T. Anwar, *"Mining feature -opinion pairs and their reliability scores from web opinion sources,"* WIMS '12, June 13-15, 2012 Craiova, Romania.

[2] Archak, Nikolay, Anindya Ghose, and Panagiotis G. Ipeirotis. *"Show me the money!: deriving the pricing power of product features by mining consumer reviews."* Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-2007). 2007.

[3] Arjun Mukherjee and Bing Liu, *"Mining Contentions from Discussions and Debates,"* KDD'12, Beijing, China, August 12–16, 2012.

[4] B. Pang, L. Lee, and S. Vaithyanathan, *"Thumbs up?:sentiment classification using machine learning techniques,"* Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol.10, pp. 79-86, 2002.

[5] Bing Liu, *"Sentiment analysis and opinion mining."* Morgan and Claypool publishers, May 2012.

[6] B. Pang and L. Lee, *"Opinion mining and sentiment analysis,"* Foundations and Trends in Information Retrieval 2(1-2), pp. 1–135, 2008.

[7] Bolanle A. Ojokoh and Olumide Kayode, *"A feature-opinion extraction approach to opinion mining,"* Journal of web engineering, vol. 11, No.1, pp.051-063, 2012.

[8] Bollen, Johan, Huina Mao, and Xiao-Jun Zeng. *"Twitter mood predicts the stock market"*. Journal of Computational Science, 2011.

[9] Dave, Kushal, Lawrence S. and Pennock D. *"Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews."* In Proceedings of the 12th International Conference on World Wide Web, WWW 2003, pp. 519-528, 2003.

[10] Das, Sanjiv and Mike Chen. *"Yahoo! for Amazon: Extracting market sentiment from stock message boards."* Proceedings of APFA-2001, 2001.

[11] Dellarocas, C., X.M. Zhang, and N.F. Awad. *"Exploring the value of online product reviews in forecasting sales: The case of motion pictures."* Journal of Interactive Marketing, 21(4): p. 23-45, 2007.

[12] Dmitry Davidov, Oren Tsur and Ari Rappoport, *"Semi-supervised recognition of sarcastic sentences in twitter and amazon,"* Proceedings of the fourteenth conference on Computational Natural Language Learning, pp. 107-116, Uppsala, Sweden, 15-16 July 2010.

[13] Ghose, Anindya, Panagiotis G. Ipeirotis, and Arun Sundararajan. *"Opinion mining using econometrics: A case study on reputation systems."* Proceedings of the Association for Computational Linguistics (ACL), 2007.

[14] Hatzivassiloglou, Vasileios and Kathleen R. McKeown. *"Predicting the semantic orientation of adjectives."* Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-1997), 1997.

[15] Hearst M., *"Direction-based text interpretation as an information access refinement in Text-Based Intelligent Systems."* P. Jacobs, Editor 1992,Lawrence Erlbaum Associates. Pp- 257-274.

[16] Hu, Nan, Paul A Pavlou, and Jennifer Zhang. *"Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication."* Proceedings of Electronic Commerce (EC), 2006.

[17] Junichi Tatemura. *"Virtual reviewers for collaborative exploration of movie reviews."* In Proceedings of Intelligent User Interfaces (IUI), pages 272–275, 2000.

[18] Loren Terveen, Will Hill, Brian Amento, David McDonald, and Josh Creter. *"PHOAKS: A system for sharing recommendations"*. Communications of the Association for Computing Machinery (CACM), 40(3):59–62, 1997.

[19] Minqing Hu and Bing Liu, *"Mining and summarizing customer reviews."* Proceedings of the 10th ACM SIGKDD International conference on knowledge discovery and data mining, 2004.

[20] Michael J Paul, ChengXiang Zhai and Roxana Girju, *"Summarizing contrastive viewpoints in opinionated text."* Proceedings of the 2010 conference on the empirical methods in natural language processing, MIT Massachusetts, pp. 66-76, October 2010.

[21] Murthy G. and Bing Liu, *"Mining opinions in comparative sentences."* Proceedings of the 22nd international conference on computational linguistics(Coling 2008), pp. 241-248, Manchester, August 2008.

[22] Nasukawa, Tetsuya and Jeonghee Yi, *"Sentiment analysis: capturing*

favourability using natural language processing." Proceedings of the K-CAP03, 2nd International Conference on knowledge capture, 2003.

[23] N. Jindal and B. Liu, *"Review spam detection."* Proceedings of WWW 2007, Banff, Alberta, Canada, May8-12, 2007. (Poster paper).

[24] N. Jindal and B. Liu, *"Opinion spam and analysis."* Proceedings of the Conference on Web Search and Web Data Mining (WSDM), pp. 219–230, 2008.

[25] P.D. Turney, *"Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews."* Proceedings of the Association for Computational Linguistics (ACL), pp. 417–424, 2002.

[26] Park Do-Hyung, Jumin Lee, and Ingoo Han. *"The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement."* International Journal of Electronic Commerce,11(4), p. 125-148, 2007.

[27] Paula Carvalho, Luis Sarmento, Mario J. Silva and Eugenio de Oliveira, *"Clues for Detecting Irony in User-Generated Contents:Oh...!! It's "so easy" ;-),"* TSA'09, Hong Kong, China, November 6, 2009.

[28] Peter D. Turney and Michael L Littman, *"Measuring Praise and criticism: inference of semantic orientation from association."* ACM Transactions on Information Systems, TOIS 2003, 21(4), pp. 315-346, 2003.

[29] Riloff, E & Wiebe, J. *"Learning extraction patterns for subjective expressions,"* EMNLP'03, 2003.

[30] Roberto G. Ibanez, Smaranda Muresan and Nina Wacholder, *"Identifying Sarcasm in Twitter: A Closer Look,"* Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: short papers, pp. 581-586, Portland, Oregon, June 19-24, 2011.

[31] S. Morinaga, K. Yamanishi, K. Tateishi and T. Fukushima, *"Mining product reputations on the web,"* SIGKDD'02, Edmonton, Alberta, Canada, 2002.

[32] Tumasjan, Andranik, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe. *"Predicting elections with twitter: What 140 characters reveal about political sentiment."* Proceedings of the International Conference on Weblogs and Social Media (ICWSM-2010). 2010.

[33] Wiebe Janyce, *"Identifying subjective characters in narrative."* Proceedings of the International Conference on Computational Linguistics (COLING-1990), 1990.

[34] Wiebe Janyce, *"Tracking point of view in narrative."* Computational Linguistics, pp- 233–287, 1994.

[35] Wiebe Janyce, *"Learning Subjective Adjectives from Corpora,"* American Association for Artificial Intelli-gence (www.aaai.org), 2000.

[36] Won Y. Kim, Joon S. Ryu, Kyu Kim and Ung M. Kim, *"A method for opinion mining of product reviews using association rules,"* ICIS 2009, November 24-26, 2009 Seoul, Korea.

[37] Xiaowen Ding, Bing Liu and Philip S. Yu, *"A holistic lexicon -based approach to opinion mining."* WSDM'08, February 11-12, 2008, Palo Alto, California, USA.

[38] Xiaowen Ding and Bing Liu, *"Resolving object and attribute coreference in opinion mining."* Proceedings of 23rd international conference on computational linguistics (Coling 2010), pp. 268-276, Beijing, August 2010.

[39] Yang Liu, Xiangji Huang, Aijun An, and Xiaohui Yu. *"ARSA: A sentiment - aware model for predicting sales performance using blogs."* Proceedings of ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR-2007). 2007.

[40] Yubo Chen and Jinhong Xie, *"Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix,"* Management Science, vol 54, no 3, pp. 477-491, March 2008.

[41] Zhongwu Zhai, Bing Liu, Lei Zhang, Hua Xu and Peifa Jia, *"Identifying evaluative sentence in online discussions."* Association for the advancement of artificial intelligence (www.aaai.org), 2011.

[42] Zhongwu Zhai, Bing Liu, Hua Xu and Peifa Jia, *"Clustering product features for opinion mining,"* WSDM'11, February 9–12, 2011, Hong Kong, China.