Sentiment Analysis Techniques: A Review

Akansha Srivastava
Research Scholar
Rkdf Institute of Science and Technology
Bhopal (M.P.)

Ravindra Gupta
Assistant professor
Rkdf Institute of Science and Technology
Bhopal (M.P.)

Abstract:- **Sentiments are the attitude, opinions, thoughts, beliefs or feelings of the writer towards something, such as people, artifacts, company or location. Sentiment analysis intends to conclude the judgment of a presenter or an author apropos to some subject matter or on the whole relative polarity of the manuscript. The outlook could be the perception or assessment, emotional condition, or the projected poignant message of the person behind.**

Keywords:- Sentiment Analysis, Machine learning, Classification.

I. INTRODUCTION

Sentiments are the attitude, opinions, thoughts, beliefs or feelings of the writer towards something, such as people, artifacts, company or location. Sentiment analysis intends to conclude the judgment of a presenter or an author apropos to some subject matter or on the whole relative polarity of the manuscript. The outlook could be the perception or assessment, emotional condition, or the projected poignant message of the person behind. Opinions are decisive influencer of our behavior. Our views and insights of veracity are conditioned on how others perceive the world. The rudimentary job in opinion mining deals with deducing the inclusive polarity of the document on some specific subject matter. Sentiment analysis is a 'suitcase' field of research that contains numerous diverse disciplines, not just associated to computer science but also to social disciplines, such as psychology, philosophy, and ethics [6]. Sentence level classification involves two tasks. The purpose of primary task is to verify the nature of statement i.e. subjective or objective. Subjective means individual's own interpretation and objective opinion means that you are looking as an outsider or another person. The main aim of second task is to verify if the subjective sentence is positive, negative, or neutral. There are mainly two steps included in this process:

➢ Subjective classification of a sentence into one of two categories i.e. objective and subjective
➢ Sentiment classification of subjective sentences into two categories i.e. positive and negative

Generally, truthful information is presented by an objective sentence whereas a subjective sentence articulates individual feelings, views, sentiments, or values. There are several techniques using which subjective sentence can be identified e.g. Naïve Bayesian classifier. Nevertheless, it is merely not sufficient to know whether the sentence contain a positive or negative opinion. This is an intermediary step that provides support in filtering out sentences having no

opinions. A subjective sentence may include numerous opinions and subjective and truthful parts. There are several prime data mining strategies applied to extract facts and information. Figure 1 shows the techniques of Opinion Mining. A lot of steps are included in the whole process. These steps include online cleaning of text, removing white spaces, amplifying acronym, stemming, stop word elimination, refusal managing and lastly feature selection.

Later, opinions are classified as positive, negative and neutral using classification approaches.
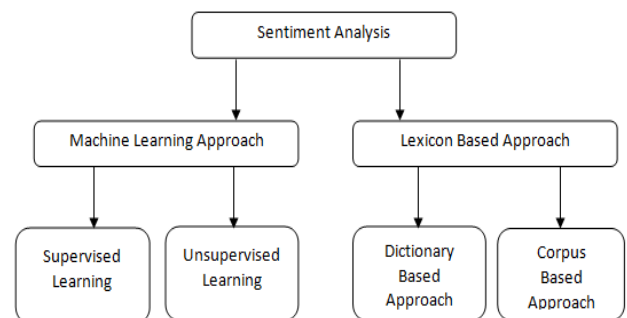

Fig 1:- Techniques of Sentiment Analysis

Machine learning is completely based on machine learning approaches. These approaches provide solution of sentence level classification issue. Also, these approaches make the decree of syntactic features. Machine learning approaches are of two types namely supervised learning and unsupervised learning. Machine learning is expected to allow machines to adjust their interior configuration in such a way that they can predict the upcoming performance boost.

➢ **Supervised Learning:** Supervised learning considers classification issues. The general purpose is to obtain the workstation to discover a classification scheme that we have formed. Digit recognition, once again, is a well known example of classification learning. More widely, classification learning is appropriate for any problem where classification learning is valuable and classification detection is easy. In some cases, it might not be compulsory to give programmed classifications to every occurrence of a problem if the method can itself perform classification.
➢ **Unsupervised Learning:** Without referring any labeled results, the patterns of any dataset are assumed through these types of algorithms. In contrast to supervised machine learning, it is not possible to apply unsupervised machine learning techniques to a regression or a classification problem. This makes the training of algorithm complicated in normal way. In its

place, unsupervised learning can be utilized for discovering the underlying data structure. Unsupervised machine learning is used to reveal earlier unknown data patterns [21].

II. LITERATURE REVIEW

Hu and Liu (2004) [7] performed opinion mining of online product reviews in 3 steps: (1) features of goods which have been remarked on by users are taken out first; (2) opinion sentences are discovered in each review and then decision is taken whether each opinion is positive or negative (3) demonstrating the result. They collected reviews of numbers of products sold online like MP3 Players, DVD's, digital camera and mobile phones from Amazon.com and CNN.net. Opinion word extraction and aggregation is the main technique used by them and features are preferred on the basis of opinion words itself. Their contribution resulted in efficient performance as compared to opinion sentences extraction for DVD-73%, and MP3-93%. The overall accuracy of five products is achieved from 64% to 84%.

Godbole et al. (2007) [13] proposed classification in a lexicon obtained from Word Net. They designed different lexicons for each topic. So, lexicon for politics is totally different from that for health. From an initial lexicon, they designed a graph model to expand polarities to other words. For instance, if the word "good" is marked as positive, all synonyms of "good" are marked as positive and all antonyms of "good" are marked as negative. Then, a new iteration is performed for next level (with the synonyms of the synonyms and the antonyms of the antonyms) and so on. Depending on the distance, the polarity score is different. Applying the formula $1/c^d$ where $c > 1$ and d is the number of nodes away. With this kind of formulation, the system ends up with polarities defined for all the words. After getting score for all words, we can calculate polarity scores of each text by dividing the sum of all polarity scores in a text between numbers of total words. The score was tested against names of celebrities i.e. Maria Sharapova got the best score.

Esuli and Sebastiani (2007) [14] presented an extremely interesting scheme that applied page rank algorithm to determine term polarities. For this purpose, they used extended WordNet to build a graph where each synset has certain polarity depending on the polarity of its members. The main hypothesis is that there won't be huge variations and each synset will have a similar degree of negativity. This will produce a graph of relation between different synsets that will transfer its polarity properties to its neighbors. One interesting point of this experiment is

that they computed the page rank separately for positive synset and negative synset (starting the entire graph from scratch for each case). Also, it is noticed that effectiveness is much better with positive terms. This means that classifying negative terms is a harder task. As a conclusion, they see that this type of model can be applied to other cases related with semantic properties of words.

K. Cai et al. (2008) [19] explained sentiment analysis which included a classification method along with an opinion based approach. The opinion classification element differentiated the comparative sentiment expressed by the terms in all fragments and then partitioned the fragments into positive, negative, and neutral groups. The sentiment subject recognition module identifies the important areas implied beyond every sentiment group by word support metrics.

M. Eirinakiet al. (2012) [33] proposed an opinion search engine scheme. The proposed approach integrated the pair of opinion mining algorithms. The outlooks are based on features and the position of these outlooks is also substantially built on the features as a substitute of an object as a whole. Inhabitants appear to dislike a precise object as of several features allied with the result. Their primary experimental assessment on numerous patron review data sets has exposed that their findings achieved extremely high level of accuracy.

Karamibekr and Ghorbani (2012) [32] firstly carried out an arithmetical exploration on the divergence among sentiment analysis of products and social issue. Then, on the basis of some conclusions, they proposed a scheme to consider the part of verb as the most imperative expression in conveying opinions concerning the societal matters. Statistical and experimental fallouts confirm that making an allowance for verbs not merely is essential and definite, other than that they also augment the concert of sentiments analysis. They collected their data from Procon.org, yahoo and CNN answers. Features are picked on the hinge of opinion directories and opinion structure. Formed on verb-oriented method result are calculated as 65% for social issues and 62.5% for car models.

K. Ghag and K. Shah (2013) [35] surveyed that Sentiment Analyzers are based on language. Various practices used a dictionary to collect opinion. Few techniques used training set while others used both training set and dictionary. No existing method is widespread sufficiently to be language independent. This clearly stated the necessity of hard work to demonstrate Sentiment Analyzer without utilizing training dataset.

Author	Year	Description	Outcome
Hu and Liu	**2004**	They collected reviews of numbers of products sold online like MP3 Players, DVD's, digital camera and mobile phones from Amazon.com and CNN.net. Opinion word extraction and aggregation is the main technique used by them and features are preferred on the basis of opinion words itself.	Their contribution resulted in efficient performance as compared to opinion sentences extraction for DVD-73%, and MP3-93%. The overall accuracy of five products is achieved from 64% to 84%.
Godbole et al	**2007**	They designed different lexicons for each topic. So, lexicon for politics is totally different from that for health. From an initial lexicon, they designed a graph model to expand polarities to other words.	After getting score for all words, we can calculate polarity scores of each text by dividing the sum of all polarity scores in a text between numbers of total words. The score was tested against names of celebrities i.e. Maria Sharapova got the best score.
Esuli and Sebastiani	**2007**	For this purpose, they used extended WordNet to build a graph where each synset has certain polarity depending on the polarity of its members. The main hypothesis is that there won't be huge variations and each synset will have a similar degree of negativity. This will produce a graph of relation between different synsets that will transfer its polarity properties to its neighbors.	This means that classifying negative terms is a harder task. As a conclusion, they see that this type of model can be applied to other cases related with semantic properties of words.
K. Cai	**2008**	The opinion classification element differentiated the comparative sentiment expressed by the terms in all fragments and then partitioned the fragments into positive, negative, and neutral groups	The sentiment subject recognition module identifies the important areas implied beyond every sentiment group by word support metrics.
M. Eirinakiet	**2012**	The proposed approach integrated the pair of opinion mining algorithms. The outlooks are based on features and the position of these outlooks is also substantially built on the features as a substitute of an object as a whole. Inhabitants appear to dislike a precise object as of several features allied with the result.	Their primary experimental assessment on numerous patron review data sets has exposed that their findings achieved extremely high level of accuracy.

Table 1:- Table of Comparison

III. CONCLUSION

Sentiment analysis is a 'suitcase' field of research that contains numerous diverse disciplines, not just associated to computer science but also to communal disciplines, such as psychology, philosophy, and ethics. The sentiment analysis methods which are proposed so far have various steps. In the pre-processing stage, the missing and redundant values are removed from the dataset. The feature extraction method established relationship between attribute and target set. In the last step of classification, the classification method is enforced which can categorize data into certain classes like positive, negative and neutral.

REFERENCES

[1]. J. M. Wiebe, R. F. Bruce, and T. P. O'Hara, "Development and use of a Gold-standard Data Set for Subjectivity Classification." Proceeding of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, USA, pp. 246-253, 1999.

[2]. D. Pelleg and A. Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," in Proc. of the 17th Int. Conference on Machine Learning, San Francisco, USA, pp. 727-734, 2000.

[3]. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," Conference on Empirical Methods in Natural Language Processing, USA, pp. 79-86,2002.

[4]. E. Riloff and J. Wiebs, "Learning Extraction Patterns for Subjective Expressions," Conference on Empirical Methods in Natural Language Processing, Japan, pp.105-112, 2003.

[5]. T. Wilson and J. Wiebe, "Annotating opinions in the world Press," 4th SIG dial Workshop on Discourse and Dialogue, Sapporo, Japan, pp. 13-22, 2003.

[6]. K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," Proceedings of WWW, 2003, pp. 519–528.

[7]. Hu and Liu, "Mining and Summarizing Customer Reviews," in International Conference on Knowledge Discovery and Data Mining, Seattle, USA, pp. 168-177, 2004.

[8]. B. Pang and L. Lee, "A sentimental education: Sentimental analysis using Subjectivity Summarization based on Minimum cuts," Proceeding of the 42nd Annual Meeting on Association for Computational Linguistics, USA, pp. 271-278, 2004.

[9]. S. M. Kim and E. Hovy, "Determining the Sentiment of Opinions." Proceedings of the 20th International Conference on Computational Linguistics, USA, pp. 1367-1373, 2004.

[10]. A. M. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews," Conference on Human Language Technology and Empirical Methods in Natural Language Processing, British Columbia, pp. 339-346, 2005.

[11]. T. Wilson, J. Wiebe, and Paul Hoffmann, "Recognizing Contextual Polarity in Phrase-level sentiment analysis," Proceedings of the conference on human language technology and empirical methods in natural language processing, USA, pp. 347-354, 2005.

[12]. M. Chau and J. Xu, "Mining communities and their Relationships in Blogs: A study of online hate groups," International Journal of Human – Computer Studies, vol. 65, issue 1, pp. 57-70, 2007.

[13]. N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-Scale Sentiment Analysis for News and Blogs," International Conference on Weblogs and social Media, USA, pp.21-24, 2007.

[14]. A. Esuli and F. Sebastiani, "PageRanking WordNet Synsets: An Application to Opinion Mining," 45th Annual Meeting-Association for Computational linguistics, Prague, Czech Republic Vol.45, pp. 424-431, 2007.

[15]. B. Pang and L. Lee, "Opinion Mining and Sentimental Analysis," Foundations and Trends in Information Retrieval, USA, vol.2, issue 1-2, pp. 1-135, 2008.

[16]. Liu B., "Opinion Mining and Summarization," World Wide Web Conference Beijing, China, 2008, Downloaded from: https://www.cs.uic.edu/~liub/FBS/opinion-mining-sentiment-analysis.pdf [21st June 2016]

[17]. P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, USA, pp. 417-424, 2008.

[18]. A. Beygelzimer, J. Langford, and B.Zadrozny, "Machine Learning Techniques- Reductions between Prediction Quality Metrics," Performance Modeling and Engineering Springer, pp. 3-28, 2008.

[19]. K Cai, S. Spangler, Y. Chen, L. Zhang, "Leveraging Sentiment Analysis for Topic Detection," International Conference on Web Intelligence and Intelligent Agent Technology, pp. 265-271, 2008.

[20]. M. Annett, G. Kondrak, "A comparison of sentiment analysis techniques: Polarizing movie Blogs", In Canadian Conference on AI, pp. 25–35, 2008.

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

[22]. A. Abbasi, H. Chen, and A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums," In ACM Transactions on Information Systems, vol. 26 Issue 3, pp. 1-34, 2008.

[23]. A. Agrawal, F. Biadsy, and K.R. Mckeown, "Contextual Phrase-Level Polarity Analysis using Lexical Affect Scoring and Syntactic n-grams," Proceeding of the 12th Conference of the European Chapter of the Association for Computational Linguistics, Athens, Greece, pp. 24-32, 2009.

[24]. Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches", Expert Systems with Applications, vol. 36, pp. 6527-6535, 2009

[25]. B. Liu, "Handbook of Natural Language Processing," Chapter Sentiment Analysis and Subjectivity, Second edition, ISBN 978-1420085921, pp. 1-38, 2010.

[26]. Bing Liu, "Sentiment Analysis: A Multi-Faceted Problem," Journal of IEEE Intelligent Systems, vol.25, issues 3, pp. 76-80, 2010.

[27]. S. Li, H. Zhang, W. Xu, G. Chen and Jun Guo, "Exploiting Combined Multi-level Model for Document Sentiment Analysis," International Conference on Pattern Recognition, pp. 4141-4144, 2010.

[28]. K. Xu, S. S. Liao, J. Li, Y. Song, "Mining comparative opinions from customer reviews for Competitive Intelligence," Decision Support Systems, vol. 50, issue 4 , pp. 743–754, 2011.

[29]. R. Lau, W. Zhang, P. Bruza and K. Wong, "Learning Domain-specific Sentiment Lexicons for Predicting Product Sales," IEEE International Conference on e-Business Engineering, pp. 131-138, 2011.

[30]. L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis", Technical report, HP Laboratories, 2011.

[31]. Ji Fang and Bi Chen, "Incorporating Lexicon Knowledge into SVM Learning to Improve Sentiment Classication", In Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), pages 94–100, 2011.

[32]. M. Karqmibekr and A.A. ghorbani, "Sentiment Analysis of a Social Issues," International Conference on a Social Informatics, USA, pp. 215-221, 2012.

[33]. M. Eirinaki, S. Pisal, J. Singh, "Feature-based opinion mining and ranking," Journal of Computer and System Sciences, vol. 78, issue 4, pp. 1175- 1184, 2012.

[34]. E. Haddia, X. Liua and Y. Shib, "The Role of Text Pre-processing in Sentiment Analysis," Information Technology and Quantitative Management, vol. 17, pp. 26-32, 2013.

[35]. K. Ghag and K. Shah, "Comparative Analysis of the Techniques for Sentiment Analysis," International Conference on Advances in Technology and Engineering (ICATE), pp. 1-7, 2013