

Comparative Analysis of Review Analysis Algorithms

Raj Kumar Saini¹, Prakash Dangi²

¹Student, Department of Computer Science and Engineering, Modern Institute of Technology & Research Centre, Alwar, India

²Assistant Professor, Department of Computer Science and Engineering, Modern Institute of Technology & Research Centre, Alwar, India

Abstract: This paper deals with the study of the papers which emphasize their research over the opinion mining and Naïve Bayes algorithms and reviews the textual content as per the trained dataset modules. With the pervasiveness of online discussion forum and social media platform, user generated text containing opinions on some hot issues has increased significantly. Due to the large amount of such emotional reviews and posts on Internet, it is impossible for users to digest such information manually. Therefore, automatically mining opinion from online texts, aiming at discovering user concerned topics and the corresponding opinion, becomes essential.

Keywords: Opinion Mining, Naïve Bayes Classifier, Data Mining, Machine Learning, Text Analysis.

1. Introduction

With the tense development of client created messages on the Internet, to reduce valuable data consequently from submitted documents get interests from analysts in different fields, specifically groups of Natural Language Processing (NLP).

Moreover, so many applications of opinion mining, like “product pricing”, “competitive intelligence”, “market prediction”, “election forecasting”, “nation relationship analysis, risk detection in banking systems” manage high focusing from industrial, manufacturing and trade areas furthermore social media is growing, online review sites, like Amazon, Twitter and other like them providing the huge corpora which are vital assets for educational research. It has both Interests to the both scholarly community and industry to improvement of opinion mining. There' sentiments can be critical when it's a great opportunity to settle on a choice or pick among numerous choices. At the point when those decisions include important assets (for instance, investing cash to purchase items or services) individuals frequently depend on their companions' past happenstances. As of not long ago, the fundamental wellsprings of data were companions and concentrated magazine or sites. Presently, the Social Networking Sites (SNS) gives new apparatuses to productively make and offer opinions with everybody associated with social networking sites via Web 2.0 enable individuals to share valuable data and opinions about the products or services

purchased online. However, the shared data and opinions are unstructured may include emotions, sentiments, characteristics, numbers, dates, and facts. These days, it is the center of attention for majority of researchers and scientists to gather and capture popular sentiments about the socials, political growths, business procedures, and the products liking or disliking from the business world. Hence, as an outcome of it, they can conceivably promote and forecast the financial share of the market for the business. The subsequent developing fields are opinion mining and sentiments analysis. In modern e-businesses, the merchants invest on opinion mining and sentiment analysis research then utilize the findings to enhance client relationship administration and suggestion frameworks through positive and negative clients' input. As a result, the findings may also help distinguish and prohibit "flares" (excessively warmed or hostile dialect) in social correspondence and upgrade antispam frameworks within the organization. More, the businesses can evaluate and foresee open states of the public thoughts towards their products and services, and reputation.

Opinion mining is a growing field to identify the thoughts and sentiments of people, which they express in form of their feedbacks or reviews on various things. Today due to vast use internet and social platforms, people are having a huge amount of space where they can publically express their opinions.

These reviews are present in various forms on web like the feedbacks for products listed on various ecommerce web sites, or the personal posts from Facebook, twitter, bloggers etc. Some formal reviews are also available in various discussion forums related to products/sites or domains. People also post a lot of personal views in form of movie reviews or the buzz creating news in various articles for magazines and newspapers. These opinions are directly related to how they feel. And this feeling can be classified as being positive, negative or neutral in nature. Positive views have a positive impact on society and negative views creates a negative impact.

Today, media also plays an important role in developing a person's views and thoughts related to any product or scenario. Any news article that a newspaper publishes sometimes creates negative while sometimes positive buzz about any scenario to

common public. This buzz directly impacts the society on a large scale.

2. Literature survey

In association rule mining, the problem with the algorithm is that, the algorithm does not consider the position of the words in the sentence. Pruning is used in this technique to remove or cut off the aspects with high frequency, compactness and redundancy. Only limitation of this technique is that the aspects found out by the mining algorithm may or may not be aspects of the product. The pruning rules cannot eliminate these false aspects. Mir, et. al. [1] presents an effective model for aspect based opinion mining which satisfies most of the crucial factors for effective opinion mining. It basically has four main phases which are Preprocessing Phase, Feature Extraction and Opinion Word Phase, Sentiment Analysis Phase and finally Result Summarization Phase. This proposed model is yet to validate its efficiency in real world datasets.

Cernian, et. al. [2] came up with a semantic approach for sentiment analysis, which uses the SentiWordNet lexical resource. SentiWordNet, a public lexical resource mainly designed for opinion mining and sentiment analysis. Three types of scores to each word is assigned by the same: positivity, negativity and objectivity. The experimental results proved a 61% average rate of success for a set of 300 product reviews from Amazon website. H. Xu, et. al. [3] proposed a method which uses explicit topic mining model to detect implicit aspects. Support vector machine classifier has been used in this model to identify implicit aspects. The problem with this model was that this method could be implemented in small dataset. The model was neither domain adaptable nor language.

Moghaddam, S, et. al. [4] proposed a method which answers the questions about products or items from reviews of customers/users. The Pattern Matching technique has been used for aspect detection and Naive Bayes, SVM & KNN algorithms are used for polarity detection. Because of some limitations of the POS tagger Q-type detection and target product extraction this proposed method has given lower correct percentage. Jingbo Zhu, et. al. [5] proposed an approach which does not include the traditional survey questionnaires and evaluates the customer's opinion from customer's reviews. The Bootstrapping-based ART learning algorithm has been used on unlabeled dataset for identification of aspects for sentiment analysis. The proposed algorithm does not work on multi-aspect sub sentences. Asghar, et. al. [6] proposed a method which extracts aspects from the user reviews by using semi-supervised technique. The mining technique that has been used is SAS Model (seeded aspect and sentiment model) and the second mining technique is ME-SAS Model (Maximum-Entropy). This technique is not fully intelligent in detecting aspects from the reviews and then grouping the related/similar terms into a single category. It requires help from the customer/ user who types the review online for initial seed. Chinsha, T. and S. Joseph [7] proposed an approach which uses a syntactic

approach for aspect based opinion mining. This opinion mining approach focusses on syntactic dependency, average score of opinion words, SentiWordNet as well as aspect table. This model ignores the issue of multi aspect, implicit aspect, comparative sentences etc. The dataset they used is very small and is not applicable for large datasets.

In [8], authors anticipated aspect based opinion mining using support vector machine classifier Hu and Liu [9, 10] presented a new technique that performs extraction and summarization of customer reviews by using association rules based on an apriori algorithm. Carenini et al. in [11] required improving the aspect extraction of prior designs using output from Hu and Liu's [10] model as input to their system to confine knowledge from customer reviews.

Wu et al. in [12] proposed a novel approach to classify noun and verb phrases as aspect/features and opinion expressions, and relationship between them. In [13], Qi and Chen proposed a discriminative model by using linear-chain conditional random files to mine opinions. Outcome of this model adjourn improvements in recall and precision compared to new methods proposed by Turney [14] and Jin et al.

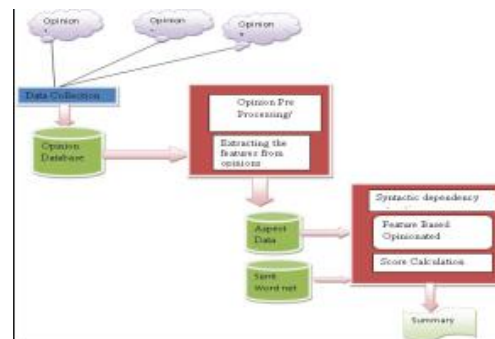


Fig. 1. Proposed architecture for opinion mining

The architecture of the proposed opinion mining method [15] is given in the Figure 1. The model is very secure and has no chances of data leakage where product opinion and rating is autonomous. There by the logs of the reviewed products and their review ratings are stored and cannot be altered in the near future.

3. Conclusion

The above studied papers has the combination of the opinion mining and the sentiment analysis as well. The studies from the past shows that the opinion mining requires the concept of Machine Learning to blend to obtain the best of the results and to ensure the reliability of the concept to be fair enough.

References

- [1] J. Mir and M. Usman, "An effective model for aspect based opinion mining for social reviews," in Digital Information Management (ICDIM), 2015 Tenth International Conference on, pp. 49-56. IEEE 2015.
- [2] A. Cernian, V. Sgarciu, and B. Martin, "Sentiment analysis from product reviews using SentiWordNet as lexical resource," in Electronics,

- Computers and Artificial Intelligence (ECAI), 2015 7th International Conference on, pp. WE-15. IEEE 2015.
- [3] H. Xu, F. Zhang and W. Wang, "Implicit feature identification in Chinese reviews using explicit topic mining model," *Knowledge-Based Systems*, vol. 76, pp.166- 175, 2015.
- [4] S. Moghaddam and M. Ester, "AQA: aspect-based opinion question answering," in *Proceeding of the IEEE 11th international Conference on Data Mining Workshops (ICDMW'15)*, pp. 89- 96, 2011.
- [5] J. Zhu, H. Wang, M. Zhu, B. K. Tsou and M. Ma, "Aspect-Based Opinion Polling from Customer Reviews," *IEEE Trans. Affective Comput.*, vol. 2, no. 1, pp. 37-49, 2011.
- [6] M. Z Asghar, A. Khan, S. Ahmad, and F. M Kundi, "A review of feature extraction in sentiment analysis," *Journal of Basic and Applied Scientific Research* 4, no. 3 (2014): 181-186.
- [7] T. Chinsha and S. Joseph, "A syntactic approach for aspect based opinion mining," in the *Proceeding of IEEE 9th International Conference on Semantic Computing (ICSC'15)*, pp. 24-31, 2015.
- [8] Varghese R and Jayasree M, "Aspect based sentiment analysis using support vector machine classifier", *Advances in Computing, Communications and Informatics (ICACCI)*, 2013 International Conference on, IEEE 2013, pp. 1581-1586.
- [9] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp.168- 177.
- [10] M. Hu and B. Liu, "Mining opinion features in customer reviews," in *AAAI*, 2004, pp. 755-760.
- [11] G. Carenini, R. T. Ng, and E. Zwart, "Extracting knowledge from evaluative text," in *Proceedings of the 3rd international conference on Knowledge capture*, 2005, pp. 11-18.
- [12] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, 2009, pp. 1533-1541.
- [13] L. Qi and L. Chen, "A linear-chain crfbased learning approach for web opinion mining," in *Web Information Systems Engineering–WISE 2010*, ed: Springer, 2010, pp. 128-141.
- [14] P. D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," 2002, pp. 417-424.
- [15] Singh, V. K., et al. Sentiment analysis of movie reviews: "A new feature-based heuristic for aspect-level sentiment classification. *Automation, Computing, Communication, Control and Compressed sensing (iMac4s)*", 2013 International Multi- Conference on. IEEE 2013.