

Opinion Mining – A Birds Eye View of Issues and Challenges

Abhaya Kumar Samal¹ Dayal Kumar Behera² Subhra Swetanisha³ Jitendra Pramanik⁴
Subhendu Pani⁵

^{1,2,3}Trident Academy of Technology, Bhubaneswar ⁴BPUT, Odisha, Bhubaneswar ⁵Orissa Engineering College, Bhubaneswar

Abstract— Opinion Mining has become important research area in e-commerce. Nowadays, people do not only navigate the web, but they also contribute contents to the Internet. Among other things, they write their thoughts and opinions in review sites, forums, social networks, and other websites. These opinions constitute a valuable resource for businesses, governments and consumers. Sentiment Analysis or Opinion Mining refers to a classification problem which deals with prediction of the polarity of words and then classifies them into positive and negative feelings with the aim of identifying attitude and opinions that are expressed in any form or language. User opinions regarding a product are generally unstructured. The emergence of various approaches and models provide a great opportunity to explore large volume of unstructured data. The objective is to give an outline of issues and approaches for mining unstructured reviews. Different application areas and challenges of mining unstructured data are also been discussed.

Key words: Opinion Mining, Sentiment Analysis, document-level Opinion Mining, Feature-level or Aspect based Opinion Mining, Opinion Lexicon

I. INTRODUCTION

This study represents a survey of recent methods or approaches and challenges related to opinion mining. Opinion Mining is commonly known as sentiment analysis [1]. This is a behavioral approach of human being to a subject of interest they encounter in their day to day life. It is an expression of human behavior involving their likes, dislikes, beliefs and perceptions of an object [15, 16] which may be a product or a service that they may intend to choose for use. Opinion Mining has now a day become the modern technique of evaluating the mood of the stake holder about a particular product or a topic. By taking recourse to sentiment analysis from several available forays, a consumer can acquire the best available product as per their requirement. Similarly, the promoters of industry, service and business [10] can keep in view the tenets of sentiment analysis improve their products and render their service and business friendly to the consumer. The role of ICT and recent growth of social media and twitters have made it easier for the stake holders to analyze the vast data on wide range of subjects and enable them to accept the one involving their required activity. This paper presents a review of various Opinion Mining approaches of unstructured texts.

The objective of this paper is to give a broad outline of various challenges, issues and to survey regarding the approaches of Opinion Mining which are organized as follows: Section II focuses on survey of general approaches for Opinion Mining. Section III furnished with various application areas where the Opinion Mining or Sentiment Analysis can be applied. Section IV furnished with different challenges associated with Opinion Mining and Section V includes conclusion.

II. RELATED WORKS

Figure 1 below shows the process of mining opinions or analyzing sentiments. In order to solve the problem of Opinion Mining, broadly two approaches have been used: document level and feature level. In this section we give brief review of document level and feature level or aspect based Sentiment Analysis.

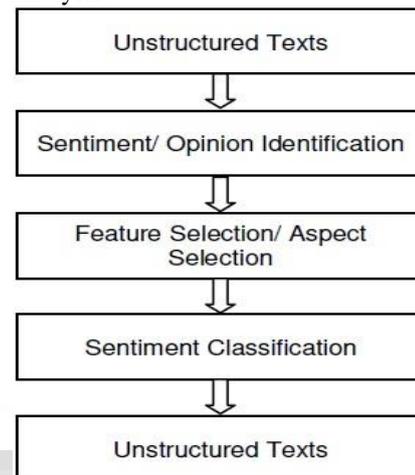


Fig. 1: Opinion Mining on Unstructured Reviews

A. Document Level Sentiment Analysis

It generates an overall opinion from one document [15]. The main aim is to classify or analyze an opinion document to produce a positive or negative opinion or sentiment.

The Document Level sentiment classification can be done by using supervised learning or unsupervised learning. As this is a text classification any supervised learning method like support vector machine (SVM), Naive Bayes classification, etc. can be applied.

P. Lee and Vaithyanathan [29] applied this approach to classify and deal with the sentiment classification problem for movie review data and classify into two classes positive and negative. It was performed quite well by using machine learning techniques like SVM or Naive Bayes and using unigrams (a bag of words).

Zhu & Goldberg[19] used a graph-based semi-supervised learning algorithm to overcome the rating-inference problem. V. Ng, S. Dasgupta and Arifin [5] classified using some linguistic knowledge sources. A. Abbasi, H. Chen and A. Salem developed a Entropy Weighted Genetic Algorithm [18] for sentiment classification in different languages. J. Kim, J.J. Li and J. H. Lee [11] and G. Paltoglou, M. Thelwall [13] used diverse information retrieval term weighting schemes and compared sentiment classification. D. Beshpalov [36] used supervised n-gram analysis for sentiment classification.

As the document level mining represent the overall opinion value, a negative sentiment or opinion regarding a product or issue does not mean that the reviewer have an

aversion to all features of the product. That is why it is important to find out which aspect the reviewer actually like and dislike. For this reason now a day's importance is given on aspect based or feature level Opinion Mining.

B. Feature Level or Aspect Based Sentiment Analysis

It gives stress on every aspect of the text. So it needs to identify each individual element which forms the opinion called as information extraction (IE) [15, 8]. There are basically two approaches followed for this: rule based & statistical method [22] and model-based method. The model-based opinion mining approach requires a training process to determining the model's parameters. After the parameter learning, it is fast to decode the new data. Jin et al. in the year 2009 demonstrated that their method is more effective and accurate in mining entities and opinions, against the rule-based method which does not have the training phase.

1) Statistical Approach

Most of the statistical approach are domain independent even language independent. Hu and Liu [28] generated product features and able to identify the opinion words which are adjacent to often appearing features. Their approach accomplished with a precision of 0.56.

E.M. Taylor, Juan D. Vel'asquez extend Bing Liu's aspect-based opinion mining technique to apply it to the tourism domain [7]. Their proposed approach is effective in finding the sentiment orientation of opinions that achieves a precision and recall of 90%.

A. Popescu & O. Etzioni Developed a system called OPINE[31] based on using statistical analysis to calculate the probability that a candidate term is relevant in a specific domain (e.g., digital cameras). Point-wise mutual information was used to measure the relevance.

C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, Herman Ng, and Chun Jin Presented a search system called Red Opal[33] that examined prior consumer reviews, identified product features, and scored each product on every feature. Which help to determine, which products to be returned, etc. – when a user specifies a desired product feature.

I. Titov, & R. McDonald proposed to adopt Latent Dirichlet Allocation (LDA) [32] which has two distinct types of topics: global topics and local topics. It first identified different domains in the corpus, and then the individual aspects of the entities in a given domain. These entity aspects serve as candidates in the opinion mining task The statistical & rule-based approach might have confined accuracy due to many entities like component, features, functions cannot be automatically identified.

2) Model Based Approach

A Model based approach using information extraction (IE) system was presented by O. Feiguina, & Lapalme. This IE discovers a model on part of speech (POS) pattern that connects the entity and aspect [34] to get the precision value of terminology extraction.

J. Kessle, & N. Nicolov used a support vector machine (SVM) for extracting feature and allied opinions. They got F-measure of 0.698 [35].

L. Yu and J. Ma, S. Tsuchiyai and F. Ren proposed a method combining HowNet and sentiment classifier [3]. This method calculates the semantic similarity of characteristic words, phrases with tagged words in HowNet.

Many works have been done using Model based approach in the literature as listed in Table 1.

Paper	Year	Technique
[29]	2002	Naive Bayes, Maximum Entropy and Support Vector Machines (SVM)
[27]	2003	Likelihood Ratio
[28]	2004	Association Mining
[24,23]	2009	Proposed opinion miner based on lexicalized Hidden Markov Model (L-HMMs)
[30]	2012	Conditional Random field(CRF)
[17]	2013	SVM, Particle swarm optimization
[21]	2014	TOM framework using hybrid classification scheme

Table 1: Document Level Mining Using Model Based Approach

III. APPLIED DISCIPLINE

Opinion mining uses Natural Language Processing (NLP), Text-mining [9], Web mining etc. to extract the information. Opinion mining has many applications.

A. Opinion Mining On Movie Review

With the growing popularity of social media reviews about anything are posted into the internet. It mainly focuses on reviews of different movies [14]. Opinion mining uses natural processing language to identify whether the movie is good or bad, which is based on message opinion. For any opinion, it is important that what other people think about a particular product. Reviews can be positive opinion or negative opinion or neutral opinion. This justification is based on the accuracy level of classification algorithm with a suitable validation process. Paper [17] uses SVM with 10-fold cross validation and confusion matrix. The hybrid Particle Swarm Optimization (PSO) is used to improve the election of best parameter in order to solve the dual optimization problem.

B. Opinion Mining In Blogs of Agriculture

With the sharing of knowledge new things are arises so now a day's blogs change the way of sharing of knowledge. In case of agriculture blogs play vital role for analyzing, creating new things about agriculture. It helps in identifying the farmer's view about agriculture; farmers can get different information which helps in decision making. With the help of intent mining and sentiment analysis we can extract text from blog for farmer's concerns [6].

C. Detecting Trend in the Web

On the web there are fusions of freely available information. It describes a framework for detection and modeling of trends on the web [26]. Here four steps are carried out: Source of documents is crawled, Search the source, Extract the information, Retrieve opinion from the web. Thus, it helps in developing a pictorial representation in the mind of readers.

D. An Application for the Argument Web

Here public construct debate, blogs, structured conversations by the use of a software tool. Argue blogging tools make online debate for public. It not only captures the online debate and argument but also link argument data. By the use of argues blogging, users can respond their opinion on the web.

E. Recommendation Systems

Recommendation System helps to classify opinions into broadly three categories like positive, negative or neutral [4]. Recommendations are determined either by explicitly conducting dialogues with online users or by analyzing existing purchasing data from a single user or a group of users. Most frequently used recommendation techniques are collaborative filtering and content based filtering [37].

F. Spam and Fake Review Detection

Due to the heavy use of internet in day to day life, there is a great possibility of increasing the spam content on the web. Spam is typically written to mislead the users. Opinion mining classifies the mail content into classes like 'spam' or 'not spam' [12] which helps the user to block spam content.

IV. CHALLENGES

Every day online social media plays a vital role where people contribute to create content, post their views, share it and many more things. So large amount of information exist in social networking which present a new set of challenges for information searching and retrieval.

As the internet and web technologies expanding every moment, researchers take keen interest for solving the problem of opinion mining [20].

Sentimental Analysis or Opinion Mining of Twitter data and other micro blogs comes across many challenges like Named entity Recognition (NER), Anaphora Resolution, Parsing, Sarcasm, Sparsity, Accuracy etc. due to quite short length and unstructured data [21].

To do research in opinion mining is challenging as it has a great impact on NLP (natural language processing) [20, 25], political science, social science, management science, economics as all of them are affected by people's opinion.

Detection of spam and spamming activities is a major challenge in order to ensure that the opinions or reviews are trusted source of valuable information.

Some major challenges related to sentiment lexicon or opinion lexicon (a list of sentiment words or opinion words [2]: normally used to represent positive or negative sentiment like good, excellent, beautiful, bad, poor etc.) are as follows:

- 1) Presence of a positive or negative sentiment word does not always represent a positive or negative polarity. Depending on various application areas, a positive or negative opinion word may have opposite orientation. e.g. "good" is usually a positive sentiment word. But it can also signify a negative sentiment like "This is not a good idea".
- 2) Not always a sentence containing a sentiment word represents any sentiment. e.g. "Whether this idea is a good idea". Neither this sentence represents a positive sentiment nor a negative sentiment.

These are the some difficult issues and challenges in order to extract the features, to determine the positive or negative polarity accurately.

In the field of natural language processing some challenges like semantic relatedness, context dependency and ambiguity, have made Opinion Mining difficult [20].

V. CONCLUSION

This paper surveyed the current approaches of opinion mining. The research in this field is in active state due to its diversity of practical application and challenging research problems. This study exploits in the Web 2.0 era social networks, blogs, web forum are the most popular employed source of opinion retrieval, opinion mining representations and related problems. Different approaches to the mining unstructured data, application areas and challenges of opinion mining are comprehensively discussed.

As the internet and web technologies expanding every moment, researchers take keen interest for solving the problem of opinion mining [20].

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers for their valuable comments and suggestions.

REFERENCES

- [1] Walla Medhat, Ahmed Hassen, and Hoda Korashy, "Sentiment analysis algorithms and applications :A survey," Ain Shams Engineering Journal, vol. 4, no. 3, pp. 1093-1113, May 2014.
- [2] Bing Liu, Sentiment Analysis and Opinion Mining.: Morgan & Claypool Publishers, 2012.
- [3] Lei Yu, Jia Ma, Seiji Tsuchiya, and Fuji Ren, "Opinion mining: A study on semantic orientation analysis for online document," in 7th World Congress on Intelligent Control and Automation, WCICA, Chongqing, 2008, pp. 4548 - 4552.
- [4] Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, pp. 1-135, 2008.
- [5] V Ng, S Dasgupta, and SMN Arifin, "Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews," in Proceedings of the COLING/ACL Main Conference Poster Sessions, 2006, pp. 611-618.
- [6] Stavros Valsamidis, Theodosios Theodosiou, Ioannis Kazanidis, and Michael Nikolaidis, "A Framework for Opinion Mining in Blogs for Agriculture," in Procedia Technology, 2013, p. Procedia Technology.
- [7] Edison Marrese-Taylor, Juan D. Vel'asquez, Felipe Bravo-Marquez, and Yutaka Matsuo, "Identifying Customer Preferences about Tourism Products using," in Procedia Computer Science 22, 2013, pp. 182 – 191.
- [8] Changqin Quan and Fuji Ren, "Unsupervised product feature extraction for feature-oriented," Information Sciences 272, pp. 16-28, 2014.
- [9] Sheng-Tun Li and Fu-Ching Tsai, "A fuzzy conceptualization model for text mining with application," Knowledge-Based Systems 39, pp. 23-33, 2013.
- [10] Magdalini Eirinaki, Shamita Pital, and Japinder Singh, "Feature-based opinion mining and ranking," Journal of Computer and System Sciences 78, pp. 1175–1184, 2012.
- [11] Jungi Kim, Jin-Ji Li, and Jong-Hyeok Lee, "Discovering the Discriminative Views: Measuring TermWeights for," in Proceedings of the 47th Annual Meeting of the ACL

- and the 4th IJCNLP of the AFNLP, Singapore, 2009, pp. 253–261.
- [12] Bing Liu, "Sentiment Analysis and Subjectivity," in *Handbook of Natural Language Processing*, N. Indurkha and F. J. Damerau, Eds., 2010.
- [13] Georgios Paltoglou and Mike Thelwall, "A study of Information Retrieval weighting schemes for sentiment analysis," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Sweden, 2010, pp. 1386–1395.
- [14] Isidro Peñalver-Martinez et al., "Feature-based opinion mining through ontologies," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5995–6008, October 2014.
- [15] Li Chen, Luole Qi, and Feng Wang, "Comparison of feature-level learning methods for mining online consumer reviews," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9588–9601, August 2012.
- [16] Kaiquan Xu, Stephen Shaoyi Liao, and Jiexun Li, "Mining comparative opinions from customer reviews for Competitive Intelligence," *Decision Support Systems*, vol. 50, no. 4, pp. 743-754, March 2011.
- [17] Abd. Samad Hasan Basaria, Burairah Hussin, I. Gede Pramudya Anantaa, and Gede Pramudya, "Opinion Mining of Movie Review using Hybrid Method of Support," *Procedia Engineering*, vol. 53, pp. 453 – 462, 2013.
- [18] Ahmed Abbasi, Hsinchun Chen, and Arab Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," *ACM Transactions on Information Systems*, vol. 26, no. 3, June 2008
- [19] Andrew B. Goldberg and Xiaojin Zhu, "Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization," in *Workshop on TextGraphs*, at HLT-NAACL, New York City, 2006, pp. 45-52.
- [20] Khairullah Khan, Baharum Baharudin, Aurnagzeb Khan, and Ashraf Ullah, "Mining opinion components from unstructured reviews: A review," *Journal of King Saud University - Computer and Information Sciences*, vol. 26, no. 3, pp. 258–275, September 2014.
- [21] Farhan Hassan Khan, Saba Bashir, and Usman Qamar, "TOM: Twitter opinion mining framework using hybrid classification scheme," *Decision Support Systems*, vol. 57, pp. 245–257, January 2014.
- [22] Guang Qiu et al., "DASA: Dissatisfaction-oriented Advertising based on Sentiment Analysis," *Expert Systems with Applications*, vol. 37, no. 9, pp. 6182–6191, September 2010.
- [23] L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Model," *IEEE ASSP magazine*, vol. 3, January 1986.
- [24] Wei Jin, Hung Hay Ho, and Rohini. K. Srihari, "OpinionMiner: a novel machine learning system for web opinion mining and extraction," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 1195-1204.
- [25] Erik Cambria, Björn Schuller, Bing Liu, Haixun Wang, and Catherine Havasi, "Knowledge-Based Approaches to Concept-Level Sentiment Analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 12 - 14, April 2013.
- [26] Jun- Ze Wang, Zheng Yan, Laurence T. Yang, and Ben-xiong Huang, "An approach to rank reviews by fusing and mining opinions based on review pertinence," *Information Fusion*, vol. 23, pp. 3–15, May 2015.
- [27] Jeonghee Yi, Tetsuya Nasukawa, Razvan Bunescu, and Wayne Niblack, "Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques," in *Proceedings of the Third IEEE International Conference on Data Mining*, 2003.
- [28] Minqing Hu and Bing 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.
- [29] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceeding EMNLP '02 Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, 2002, pp. 79-86.
- [30] Chen, L., Qi, L., & Wang, F. Comparison of feature-level learning methods for mining online consumer reviews. *Expert Systems with Applications*, 39, 9588–9601, 2012.
- [31] Ana-Maria Popescu and Oren Etzioni, "Extracting Product Features and Opinions from Reviews," in *Natural Language Processing and Text Mining*, Anne Kao and Stephen R. Poteet, Eds.: Springer London, 2007, pp. 9-28.
- [32] Ivan Titov and Ryan McDonald, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization," in *Proceedings of ACL-08: HLT Columbus, Ohio, USA*, 2008, pp. 308–316.
- [33] Christopher Scaffidi et al., "Red Opal: product-feature scoring from reviews," in *Proceedings of the 8th ACM conference on Electronic commerce*, 2007, pp. 182-191.
- [34] Olga Feiguina and Guy Lapalme, "Query-Based Summarization of Customer Reviews," in *Proceedings of the 20th conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence*, 2007, pp. 452 - 463.
- [35] Jason S. Kessler and Nicolas Nicolov, "Targeting Sentiment Expressions through Supervised Ranking of Linguistic Configurations," in *Proceedings of the Third International ICWSM Conference (2009)*, 2009.
- [36] Dmitriy Beshpalov, Bing Bai, Yanjun Qi, and Ali Shokoufandeh, "Sentiment Classification Based on Supervised Latent n-gram Analysis," in *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011, pp. 375-382.
- [37] Alexander Felfernig, Gerhard Friedrich, and Lars Schmidt-Thieme. (2007) *Recommender Systems*. IEEE Intelligent Systems.