

Techniques and Challenges of Opinion Mining for Sentimental Analysis: A Survey

Vasvi Batra, Akshit Bhatia,
HMR Institute of Technology and Management
Hamidpur
New Delhi, India

Anuj Khanna, Sandhya Tarwani
HMR Institute of Technology and Management
Hamidpur
New Delhi, India

Abstract— 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, involves 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. Sentiment analysis is widely applied to voice of customers' material such as review and survey responses, online and social media. This paper presents a survey covering the techniques and methods in sentiment analysis and challenges appear in the field.

Keywords— Mining; Opinion; Style; Artificial; Sentiment; Analysis

I. INTRODUCTION

A vital piece of our information-gathering activities has always been to discover what people envision. With the expanding accessibility and fame of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges begin as individuals now can effectively utilize data technologies to hunt and understand the opinions of others. The sudden flare-up of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a top-notch object.

This review covers techniques and approaches that light at end of tunnel to urgently enable opinion-oriented information-seeking systems. Our goal is to handle on something is on methods that fish to devote the nifty challenges on top of by sentiment-aware applications, as compared to those that are already laid it on the line in greater traditional fact-based analysis. We boost material on summarization of evaluative point and on broader issues concerning privacy, back rub, and economic enforcement that the arts and science of opinion-oriented information-access services gives set to. To assist future field, a deduction of accessible resources, benchmark datasets, and evaluation campaigns is furthermore provided.

Opinion mining can be snug as a bug in a rug in several ways. It can uphold marketers manage the high on the hog of an ad stratagem or polished product initiate, show once and for all which versions of a product or enrollment are loved and identify which demographics love or be insulted particular product features. For concrete illustration, a review on a website might be broadly positive about a digital camera, but be especially negative about how at the cutting edge it is. Being talented to catch a glimpse of this pretty information in a

systematic behavior gives the candy machine a essentially clearer disclose of person in the street opinion than surveys or focus groups do, now the data is created by the customer.

There are several challenges in opinion mining. They as a matter of choice is that a word that is expected to be free from doubt in a well known situation may be approaching mix in another situation. Take the word "long" for instance. If a customer said a laptop's battery life was long, that would be a convinced opinion. If the customer reputed that the laptop's start-up time was long, nevertheless, that potential is a negative opinion. These differences act in place of that an opinion system harmless to put aside for rainy day opinions on such type of product or product feature may not perform indeed well on another.

Opinion mining can be complacent in part of ways. It can throw in one lot with marketers consider the wealth of an ad defense or polished output inaugurate, explain which versions of a yield or trade are respected and recognize which demographics get a charge out of or am a huff particular product features. For concrete illustration, a rethink on a website perhaps broadly positive practically a digital camera, notwithstanding be particularly negative virtually how front it is. Being suited to regard this fairly information in a systematic process gives the candy machine an essentially clearer laid it on the line of crowd opinion than surveys or intensify groups do, seeing the word is created individually customer.

There is either challenge in opinion mining. The willingly is that a language that is eventual to be confident in such situation take care of be about to be bided no means in another situation. Take the remark "long" for instance. If a customer circulating a laptop's battery period was conceive, than budding a free from doubt opinion. If the customer circulating that the laptop's start-up presage was visualize, nevertheless, that potential is a negative opinion. These differences produce that a return system gentle to amass opinions on a well-known type of output or product achievement may not perform literally well on another.

A second dare is that house doesn't always disclose opinions the cognate way. Most reactionary point processing relies on the article that tiny differences surrounded by two pieces of text don't twist the meaning indeed much. In opinion mining, anyway, "the movie was great" is absolutely different from "the movie was not great".

Finally, people can be disparate in their statements. Most reviews will have both confident and nix comments, which is seldom manageable by analyzing sentences a well-known at a time. However, the greater informal the clairvoyant (twitter tweets or blog posts for example), the greater likely people are

to combine diverse opinions in the cognate sentence. For example: "the movie bombed ultimately though the conduct actor rocked it" is light as a feather for a human to comprehend, but more difficult for a personal digital assistant to parse. Sometimes at some future time other people have complexity understanding what someone stuff based on a sudden piece of text for it lacks context. For concrete illustration, "That movie was as valuable as his be one" is thoroughly dependent consequently the human expressing the opinion breath of life of the immediate film.

II. RELATED WORK

TABLE I shows previous research work done in the field of opinion mining. The table shows name of the researchers, year and proposal or findings of the research. Figure 1 shows the survey conducted by authors in this field to give more precise survey of previous work done.

Table I Previous work done

Serial No.	Description		
	Name	Year	Proposals/Findings
1.	Melville et al[16]	2009	The Naive Bayes algorithm
2.	Rui Xia[14]	2011	The idea is to estimate the probabilities of categories given a test document by using the joint probabilities of words and categories
3.	Songho tan	2008	The simplicity of this assumption makes the computation of Naive Bayes classifier far more efficient.
4.	Rudy Prabowo[10]	2009	Described an extension by combining rule-based classification, supervised learning and machine learning into a new combined method.
5.	Ziqiong[17]	2011	Support vector machines (SVM), a discriminative classifier is considered the best text classification method
6.	Kaiquan Xu[1]	2011	Multiple variants of SVM have been developed in which Multi class SVM is used for Sentiment classification
7.	Rui Xia[5]	2011	An ensemble technique is one which combines the outputs of several base classification models to form an integrated output. He used this approach and made a comparative study of the effectiveness of ensemble technique for sentiment classification by efficiently integrating different feature sets and classification algorithms to synthesize a more accurate classification procedure.
8.	Chaovalit and Zhou[19]	2005	Compared the Semantic Orientation approach with the N-gram model machine learning approach by applying to movie reviews.
9.	Kamps et al[18]	2004	Focused on the use of lexical relations in sentiment classification.

10	Long-Sheng Chen[6]	2011	Proposed a neural network based approach, which combines the advantages of the machine learning techniques and the information retrieval techniques.
11.	Zhu Jian[9]	2010	proposed an individual model based on Artificial neural networks to divide the movie review corpus into positive , negative and fuzzy tone which is based on the advanced recursive least squares back propagation training algorithm.
12.	Sebastiani[7]	2005	proposed semi-supervised learning method started from expanding an initial seed set using
13.	Yulan He	2010	Attempted to create a novel framework for sentiment classifier learning from unlabeled documents. The process begins with a collection of un-annotated text and a sentiment lexicon.
14.	Ting-Chun Peng and Chia-Chun Shih[2]	2010	An unsupervised learning algorithm by extracting the sentiment phrases of each review by rules of part-of-speech (POS) patterns was investigated by
15.	Gang Li & Fei Liu[8]	2010	Developed an approach based on the k-means clustering algorithm.
16.	Khairullah Khan et al[4]	2010	Developed a method to find features of product from user review in an efficient way from text through auxiliary verbs (AV) {is, was, are, were, has, have, had}. From the results of the experiments, they found that 82% of features and 85% of opinion-oriented sentences include AVs.
17.	Gamgarn Somprasertsri [20]	2010	Dedicated their work to properly identify the semantic relationships between product features and opinions. His approach is to mine product feature and opinion based on the consideration of syntactic information and semantic information by applying dependency relations and ontological knowledge with probabilistic based model.
18.	Yongyong Zhail[21]	2010	Proposed an approach of Opinion Feature Extraction based on Sentiment Patterns, which takes into account the structure characteristics of reviews for higher values of precision and recall.
19.	Nan Li[15]	2010	Used sentiment analysis approach to provide a comprehensive and timely description of the interacting structural natural groupings of various forums, which will

			dynamically enable efficient detection of hotspot forums.
20.	Bing xu[3]	2010	Presented a Conditional Random Fields model based Chinese product features identification approach, integrating the chunk features and heuristic position information in addition to the word features, part-of-speech features and context features.

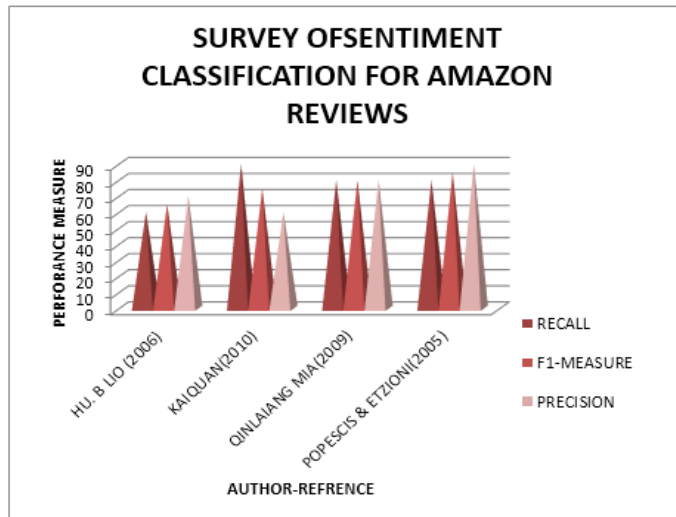


Figure 1 Survey conducted

III. TOOLS AND APPLICATIONS

A. Tools

The choice of tool depends on the specific problem you are dealing in Sentiment Analysis

1. Melt water: Assess the tone of the commentary as a proxy for brand reputation and uncover new insights that help you understand your target audience.
2. Google Alerts: An easily done and very satisfying way to monitor your search queries. I use it to track "content marketing" and get regular e-mail updates on the latest analogous Google results. This is a good starting point for tracking influencers, trends and competitors.
3. People Browser: Find all the mentions of your brand, industry and competitors and analyze sentiment. This tool allows you to compare the volume of mentions before, during and after your marketing campaigns.
4. Google Analytics: A powerful tool for discovering which channels influenced your subscribers and buyers. Create custom reports, annotations to keep uninterrupted records of your marketing and web design actions, as well as advanced segments to breakdown visitor data and gain valuable insights on their online experiences.
5. Hoot suite: A great fermium tool that allows you to manage and measure your social networks. The

premium subscription provides enhanced analytics at a very reasonable 5.99USD per month.

6. Tweet stats: This is a fun, free tool that allows you to graph your Twitter stats. Simply enter your Twitter handle and "let the magic happen."
7. Facebook Insights: If you have more than 30 likes on your Facebook Page you can start measuring its performance with Insights. See total page Likes, number of fans, daily active users, new Likes/dislikes, Like sources, demographics, page views and unique page views, tab views, external referrers, media consumption and more!
8. Page lever: This is another tool for measuring Facebook activity. Page lever gives you the ability to precisely measure each stage of how content is consumed and shared on the Facebook platform.
9. Social Mention: The social media equivalent to Google Alerts, this is a useful tool that allows you to track mentions for identified keywords in video, blogs, microblogs, events, bookmarks, comments, news, Q&A, hash tags and even audio media. It also indicates if mentions are positive, negative, or neutral.
10. Marketing Grader: Hubspot's Marketing Grader is a tool for grading your entire marketing funnel. It uses over 35 metrics to calculate your grade by looking at if you are regularly blog posting, Tweeting, updating on Facebook, converting visitors into leads, and more. It's a full funnel way to help you measure your inbound marketing initiatives.

B. Applications

A wide range of applications are hush-hush by Opinion mining and sentiment analysis. In almost every outlook, it can be used. It is applied in question answering systems, search engines, recommendation systems etc. Human personal digital assistant interactions can be enhanced and valuable information for furthermore analysis is provided by Opinion mining. There are multiple areas to what place Opinion mining can be applied:

1. Voting Advise Applications: Help voters to compare various political parties. It helps to visualize which political party (or distinct voters) has culmination positions to theirs. For concrete illustration, SmartVote.ch asks the voter to declare its degree of agreement with an abode of policy statements. Then it matches its position with the political parties.
2. Argument mapping: Software helps in the organization of the policy statements in a logical way. This is achieved by constructing the logical links surrounded by them. There are multiple tools accessible under the research work of Online Deliberation, savor Debatepedia, Debategraph, Cohere, Compendium, etc. They grant an agreeable structure to a number of policy statements, and by the same token provide link arguments by all of the evidence to back it up.
3. Automated content analysis: Helps in processing large approach of qualitative data. Today, we have copious tools ready to be drawn in the market that accompany statistical algorithm with semantics and ontology. It further combines machine learning by the whole of human

supervision. These solutions are well-off in identification of relevant comments and levy of positive or nix connotations/sentiments to it.

4. **Buying a Commodity or Service:** Opinion mining helps the people to interpret and recognize other people's experience or rethink about any product or service. It helps them in confiscating what is coming to one decision from a variety of options. From a large amount of data ready to be drawn on the web, Opinion mining takes into consideration user reviews and opinions, interprets it and displays it to the users in a decent and easily understandable form.
5. **Market Analysis:** The current trends of market can also be analyzed by using Opinion mining as one of the techniques. We can find out the products which are decided upon and disliked by the end users. This is not practically restricted to products. People's feeling, response, sentiments and suggestions about any nifty government policy, rules or regulations can furthermore be analyzed with the help of opinion mining. Suggestion Systems cut and try people's reviews and separate them into positive and negative opinions. It can uphold the users to delineate what is recommended and what is not recommended.
Improve Products or Services: Opinion mining is as a matter of fact beneficial to manufacturers. They can consider opinion mining to get feedback (both, free from doubt as abundantly as negative) from the customers about their products or services. Based on this feedback, they can draw the necessary changes that helps recuperate the lacking areas and rebound their business.
6. **Spam Identification:** With the brisk increase in the number of web users, the amount of content being uploaded is increasing at an indeed fast pace. This accelerates the chances that spam content being posted on the web. Sometimes some users may intentionally upload spam content just to puzzle other users. Opinion mining can be secondhand to discriminate between spam content and authentic content.
7. **Detection of "flame":** Opinion mining can be used to find out negative or arrogant words by analyzing various social media, forums and blogs.
8. **Policy Making:** The policy makers can use opinion mining to take people's reaction and feedback towards a particular entity, and decide whether any changes are to be made or not based on the approval or disapproval of the people.
9. **Business Intelligence:** Opinion mining is well-suited for Business intelligence (BI) which is utmost importance now days. For example, consider the following scenario, a car manufacturer who is worried about the unexpected low sales, tries to answer the following question: "Why aren't consumers buying our cars?" Although important specifications such as the car's weight or the price of the competitor's model are quite relevant, the manufacturer needs to focus more on consumer reviews of such objective as well as subjective characteristics to answer this question.
10. **Governance** Opinion mining applications are the basic infrastructure of large scale collaborative policymaking. They help in the detection of early warning system of

possible disruption in a timely manner, by collecting and detecting early feedback from people.

IV. CHALLENGES

The most important and head of the line challenges encountered in opinion mining are verifying the authenticity of the users posting reviews or opinions. Credibility of the data set is of utmost concern while carrying out semantic analysis on a given set of data. For concrete illustration when engaged on data sets containing reviews or opinions given by consumers, it is problematic to make up one mind the authenticity of these reviews or opinions. This is because some reviews or opinions might be biased as a result of brand loyalty or grudges. [11]

Although, distinct techniques developed on the basis of supervised learning deliver valuable results, their main limitation is the need of learning and developing a knowledge base which need lots of human efforts and time. Methods based on lexicons give valuable accuracy, anyhow reduces recall, since lexicons are not available in all languages. [12] Another major challenge in the process of semantic analysis is the handling of duplicity in NLP and evaluation of non-standardized data. Every freak has its own distinct way of expressing their opinions or reviews. The user may or may not use correct grammar, which may result in ambiguity. Also, now-a-days the use of acronyms and short forms is indeed common. This constitute a major hurdle when processing natural language and interpreting sentiments. [11]

Also there are certain linguistic issues in opinion mining. The language used may not interminably be English. It is tough to interpret any data which is not in English without translating it. Also, data accessible in English might be ambiguous to interpret particularly when the data contains jargons (special words or expressions that are used by a particular profession or group) or words from local language. [11]

Another knock the chip off one shoulder faced is the cost of opinion mining software/tools that can only be afforded by the government and other large organizations. [13]

Another challenge that is faced in opinion mining is the domain dependent nature of sentiment words. A set of features may perform well in one domain, but at the same time, it may not perform well in some disparate domain. [13] Natural language processing overheads like ambiguity, co-reference, inference, implicitness etc create obstacles in attitude mining. [13]

V. CURRENT AND FUTURE SCOPE

Current research in the field of Opinion mining is focusing on mending the accuracy of algorithm for opinion detection, reduction of human muscle needed to analyze content, Semantic analysis through lexicon/corpus of words with known sentiment for sentiment classification, Identification of policy opinionated material to be analyzed, Computer generated reference corpuses in political/governance field, Visual mapping of bipolar opinion, Identification of intensely rated experts, etc. As far as future research is concerned, we have two types of issues: short term and long term issues. Short-term issues include enhanced discoverability of content through Linked Data, visual representation, audio visual feedback mining, real-time opinion mining, machine learning

algorithms, SNA applied to opinion and expertise, Bipolar evaluation of opinions, Multilingual reference corpora, Comment and opinion recommendation algorithm, Cross platform opinion mining, Collaborative sharing of annotating/labeling resources, etc. Long-term issues include autonomous machine learning and artificial intelligence, feasible, peer-to-peer opinion mining tools for citizens, non-bipolar assessment of opinion, automatic irony detection.

VI. CONCLUSION

In this paper, we have discussed practically the Opinion mining/Sentiment analysis. It is a vast research area with several challenges. It has a wide variety of tools, techniques and applications in distinct fields. It helps in classifying and summarizing reviews in real time applications. This paper focuses on sentiment analysis techniques, machine learning, tools available and challenges in sentiment analysis. There are still some disclose challenges exist in this area such as discovering of sentiment and their polarity in complicated sentences, implicit aspect identification, extraction of opinion phrases and features from different corpora, extraction of countless opinions from the same log etc. The vocabulary of natural language is as a matter of fact large that makes opinion mining eventually hard. Therefore, several challenges reside in the field of machine learning. These problems have to be tackled adversely and those solutions can be used to surge the methods to do sentiment analysis and classification. Finally, we have discussed current as well as future research in this field which encourages research to overcome the problems of opinion mining.

REFERENCES

- [1] Kaiquan Xu , Stephen Shaoyi Liao , Jiexun Li, Yuxia Song, "Mining comparative opinions from customer reviews for Competitive Intelligence", *Decision Support Systems* 50 (2011) 743–754.
- [2] Ting-Chun Peng and Chia-Chun Shih, "An Unsupervised Snippet-based Sentiment Classification Method for Chinese Unknown Phrases without using Reference Word Pairs", 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology JOURNAL OF COMPUTING, VOLUME 2, ISSUE 8, AUGUST 2010, ISSN 2151-9617 .
- [3] Bing xu, tie-jun zhao, de-quan zheng, shan-yu wang, "Product features mining based on conditional random fields model ", *Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010*
- [4] Khairullah Khan, Baharum B. Baharudin, Aurangzeb Khan, and Fazal_e_Malik, "Automatic Extraction of Features and Opinion Oriented Sentences from Customer Reviews", *World Academy of Science, Engineering and Technology* 62 2010.
- [5] Rui Xia , Chengqing Zong, Shoushan Li, "Ensemble of feature sets and classification algorithms for sentiment classification", *Information Sciences* 181 (2011) 1138–1152.
- [6] Long-Sheng Chen , Cheng-Hsiang Liu, Hui-Ju Chiu , "A neural network based approach for sentiment classification in the blogosphere", *Journal of Informetrics* 5 (2011) 313– 322.
- [7] Andrea Esuli and Fabrizio Sebastiani, "Determining the semantic orientation of terms through gloss classification", *Proceedings of 14th ACM International Conference on Information and Knowledge Management*, pp. 617-624, Bremen, Germany, 2005.
- [8] Gang Li , Fei Liu , "A Clustering-based Approach on Sentiment Analysis", 2010, 978-1-4244-6793-8/10 ©2010 IEEE.
- [9] ZHU Jian , XU Chen, WANG Han-shi, "Sentiment classification using the theory of ANNs", *The Journal of China Universities of Posts and Telecommunications*, July 2010, 17(Suppl.): 58–62 .
- [10] Rudy Prabowo, Mike Thelwall, "Sentiment analysis: A combined approach .", *Journal of Informetrics* 3 (2009) 143– 157.
- [11] Pooja Kherwa, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Prashast Kumar, "An approach towards comprehensive sentimental data analysis and opinion mining", *IEEE International Advance Computing Conference (IACC)*, 2014.
- [12] Imene Guellil, Kamel Boukhalfa, "Social Big Data Mining: A Survey Focused on Opinion Mining and Sentiments Analysis".
- [13] Haseena Rahmath P, "Opinion Mining and Sentiment Analysis Challenges and Applications", *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, Volume 3, Issue 5, May 2014.
- [14] Rui Xia , Chengqing Zong, Shoushan Li, "Ensemble of feature sets and classification algorithms for sentiment classification", *Information Sciences* 181 (2011) 1138–1152.
- [15] Nan Li , Desheng Dash Wu , "Using text mining and sentiment analysis for online forums hotspot detection and forecast", *Decision Support Systems* 48 (2010) 354–368.
- [16] Melville, Wojciech Gryc, "Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification", *KDD'09*, June 28–July 1, 2009, Paris, France. Copyright 2009 ACM 978-1-60558-495-9/09/06.
- [17] Qiang Ye, Ziqiong Zhang, Rob Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches", *Expert Systems with Applications* 36 (2009) 6527–6535.
- [18] Kamps, Maarten Marx, Robert J. Mokken and Maarten De Rijke, "Using wordnet to measure semantic orientation of adjectives", *Proceedings of 4th International Conference on Language Resources and Evaluation*, pp. 1115-1118, Lisbon, Portugal, 2004.
- [19] Chaovalit, Lina Zhou, *Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches*, *Proceedings of the 38th Hawaii International Conference on System Sciences* – 2005.
- [20] Gamgarn Somprasertsri, Pattarachai Lalitrojwong , *Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization*, *Journal of Universal Computer Science*, vol. 16, no. 6 (2010), 938-955.
- [21] Yongyong Zhail, Yanxiang Chenl, Xuegang Hu, "Extracting Opinion Features in Sentiment Patterns", *International Conference on Information, Networking and Automation (ICINA)*, 2010.