

Analyzing Sentiments for Generating Opinions (ASGO)-A New Approach

Sapna Zol* and Preeti Mulay

Symbiosis Institute of Technology, A Constituent of Symbiosis International University, Pune - 412115, Maharashtra, India; sapana.zol@sitpune.edu.in

Abstract

The boom of social media over the web has great impact on individual's and organization's decision making process about certain content. In today's world of business and marketing opinion mining plays a vital role in success of launching a new product and determining which versions of product is popular within customers. To extract the purposeful content from the social web, opinion mining performs an important role. Sentiment analysis is a current topic of research in this present era. This paper depicts a small effort explaining the process, applications and some challenges of opinion mining. In addition to the three basic opinions of sentiment detection i.e. positive neutral and negative this review paper suggests more precise sentiments like exciting, irrelevant and objectionable. Overall the study interprets the initial results of the undergoing master's thesis work of integrating the sentiment and spam detection in one system.

Keywords: Exciting and Irrelevant Review Detection, Market Intelligence, Opinion Detection, Potential Customers, Sentiment Analysis

1. Introduction

Nowadays there is a huge impact of social media in everyone's life. Social media has become a platform for people to share their opinions and reviews. Blogs, forums, newspaper, micro blogs and facebook are main sources to get public review and opinion information about products. The information may consist of fact and opinion which can be extracted using Natural Language Processing (NLP) in some kind of opinionated views. To analyze the huge amount of opinionated text is a great challenge for a researcher. This domain is called as sentiment analysis also known as opinion mining. Sentiment analysis is the automated mining of opinions, emotions from sources such as text, speech, database using Natural Language Processing. No one is concerned about the detail topic when an opinion or sentiment is considered. Given a text, opinion mining systems analyzed should include: Who wrote opinion? Which part of text is opinion expressing? Is that opinion a spam?

Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". It is often referred to as subjectivity analysis, opinion mining, and appraisal extraction¹. Application of sentiment analysis is found in almost every field. Opinion mining is found to be useful for both business organization as well as individuals. When an individual wants to buy a product he will obvious go for the opinion and review of the product in the same way the business organization also wants to know the public opinion about their product. The objective of this paper is to explore the techniques and applications of opinion mining and to discover the challenges in this field. This paper is organized as follows: Section 2 describes the research work till date on sentiment and opinion mining. Section 3 describes the sentiment analysis process. Section 4 describes the applications of sentiment analysis in various fields and Section 5 summarizes the sentiment analysis study and 6 gives the partial result of the work done based on the survey.

*Author for correspondence

2. Sentiment Mining

Sentiment analysis tasks and levels are the gist of sentiment mining. Just to prove the above statement following points are considered in this paper.

2.1 Sentiment Analysis Task

Yongzheng et al². have discussed that Sentiment analysis task is mainly divided into

- Sentiment identification - identify whether a piece of text expresses opinions.
- Sentiment orientation classification - determine the orientation of opinionated text

2.2 Sentiment Analysis Level

Sentiment analysis can be done at three levels document level, sentence level, entity and aspect level. Document level expresses opinion as positive, negative of a document as a whole. Sentence level identifies if sentence is opinionated². An aspect-based opinion polling system takes as input a set of textual reviews and some predefined aspects, and identifies the polarity of each aspect from each review to produce an opinion poll^{2,3}. For example it extracts the subject of the opinion and opinion orientation. In a product review an aspect based opinion polling system will extract the subject of the opinion for example "This car is fuel efficient" So it will identify "car" (subject) and determines whether the opinion is positive or negative in this case positive.

2.3 Sentiment Analysis Process

To understand the complete flow of sentiment analysis below points are explained with the diagram Figure 1 shows the process of opinion mining

3. Data Sources

Data is collected from various social media blogs, micro blogs, review sites, and also datasets which are available online as they provide a rich level of understanding of user's opinion.

3.1 Blogs

Blogs are considered as a source or a way to spread, share, or break the information with large audience⁴. Marketing through social media is very beneficial for persons who want to get their business at top level.

Blogs are used as a source of opinion in many studies related to sentiment analysis⁴⁻⁷.

3.2 Newspapers

Newspaper is a traditional but effective tool for individuals to share their views and thoughts as well as for business organization for marketing of their products. The opinion mining study of Diana C. Mutz et al⁸. analyzed a newspaper's attempt to move community opinion and bring about policy change (Figure 1).

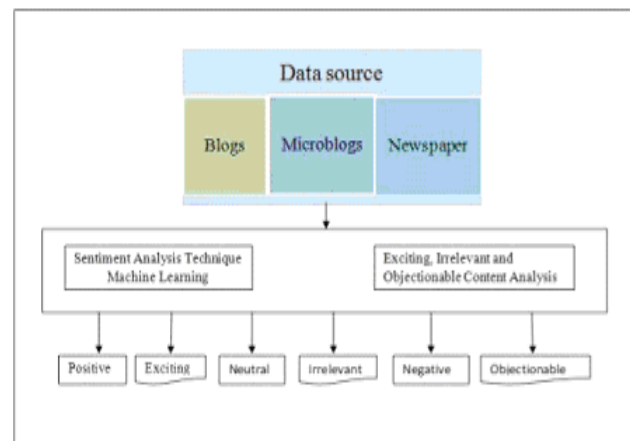


Figure 1. Opinion Mining process.

3.1.2 Sentiment Mining Techniques

Opinion analysis can be implemented by using both approaches supervised approach and unsupervised approach.

3.1.2.1 Supervised Approach

In this approach training data contains the input as well as the desired result.

3.1.2.2 Unsupervised Approach

In this approach training data does not contain the desired result.

4. Machine Learning

BO Pang et al¹⁴ was the first to use the standard machine learning algorithms Naive Bayes, maximum entropy and support vector machine for classifying the movie reviews as positive or negative. The other machine learning methods include N-gram model, K-Nearest neighborhood, ID3, C5.

4.1 Naive Bayes

Naive Bayes is mostly used for document level classification^{15,16}. A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong independence assumptions. The supervised learning method Naive Bayes uses the Bayes theorem

Where $P(A/B)$ is the conditional probability of A given B, A is the class value and B is the text we want to classify. It assumes each feature is conditional independent to other features given the class. The main advantage of this method is that a small training dataset is enough to train the classifier.

4.2 Support Vector Machine

Another supervised classifier Support Vector Machine (SVM) is found to be more effective text categorization method than Naive Bayes¹⁷. SVM is large margin classifier than just probabilistic classifiers like Naive Bayes and maximum entropy. It requires a large dataset in order to build a super quality model for classification. SVM has been used in many research studies¹⁸⁻²⁰. Like other sentiment mining applications the key for sentiment classification is effective features. Some features studied are as follows Terms and their frequency - Features like unigrams and their n-grams with their frequency count are included. Pang et al¹⁴. studied various the unigrams and bigrams feature and concluded that unigram feature give better results than the bigram or any other complicated feature in contrast Dave et al²¹. claims that bigram features outperforms unigrams in his product review. Part of speech - Part of speech tags serve as crude for word sense disambiguation²². With part of speech tags Pang et al. stated that the accuracy in results gradually increases in Naive Bayes and decreases with SVM and remains unchanged for MaxEnt. Role of negation - Negation is also necessary to take into consideration as an important feature. Not only words like 'not', 'never' expresses negation but sentence like 'The car is less fuel efficient' also conveys a negative opinion. Wilson et al²⁵. applied advanced negation modeling on expression -level polarity classification. Kennedy and Inkpen²³ and Polanyi et al²⁴. also evaluated negation models. Exciting words and abbreviations - Opinions containing words like 'OMG', 'BRB', 'WTF' should be tokenized individually. Efthymios Kouloumpis et al²⁵. tried normalization techniques such that the abbreviated words are replaced by their actual meaning.

4.3 Semantic Orientation - An Unsupervised Learning Method

Semantic orientation is a unsupervised learning method because it does not require any training dataset for mining data. Semantic orientation deals with the association between the phrases (e.g., "beautiful place", "terrible event") in the review and measures how much a word is inclined towards positive or negative. Turney²⁶ evaluated semantic orientation by using PMI (point mutual information) algorithm and accuracy of 74% is achieved

4.4 Pseudo Reviews Detection

Very few research studies give a broad idea about opinion spam in reviews²⁷⁻⁸. Pseudo reviews also called as social spam is unwanted spam content appearing on social networks and any website with user-generated content (comments, chat, etc.) (Source: Wikipedia). The very first study on opinion spam in reviews was published in 2007 and 2008 by Nitin Jindal and Bing Liu²⁷. They have classified spam reviews into three types untruthful opinion, reviews on brand only i.e. reviews not on the specific product but the brand which the product belongs and non-reviews - irrelevant reviews. They have applied logistic regression algorithm which yields good results for type 2 and type 3 spam reviews.¹⁴⁻¹⁸ are the other research done on spam detection in social media.

5. Relevance of Opinion Mining in Various Fields

Opinion Mining covers ample area of applications.

5.1 Market Intelligence

Market Intelligence can be defined as the process of acquiring and analyzing information in order to understand the market (source: mi.agri.net). Sentiment analysis discover a major role in market intelligence such that tracking customer as well as potential consumer opinions and ratings on the product, comparing and monitoring the performance etc. Most of the research work focus on sentiment and opinion mining of products e.g. movies, hotels^{14,23,29}.

5.2 Conception of Potential Customers

Potential customer is someone who is likely to purchase the product from the company in future.

stoughts of potential customers strictly depend on the opinions of current customers about the certain product. So understanding the 'would be' customers is beneficial for the success of the organization. Research Studies^{30,31} show predicting and managing behavior of creative and potential customers.

5.3 Government Sector

Use of sentiment analysis allows automatic analysis of public opinions and also predicting the results of elections^{32,33}. Opinion mining clearly identifies the firmness and weakness of government campaign³⁴.

Sentiment analysis can be applied to various domains to classify and summarize the reviews. This paper summarizes few techniques and commonly used applications of opinion mining. Further the paper describes the detail process of opinion mining. It is found that Support Vector Machine (SVM) is found to be more accurate than any other sentiment classification technique but still having some domain limitations. To boost the performance of sentiment analysis performance different methods need to be combined so that their sole limitations will be overcome. Despite the fact that the methods and techniques for sentiment classification are consistently improving still several problems in this field still remain unsolved.

6	Did plenty of research here and elsewhere before I ended up getting my Escape. ASS vehicle! Do yourself a favor, get a KandN air filter, learn to feather the gas from stops and enjoy 23+ MPG in the city (with the V6 no less!) I get 26-28 Hwy. Use the cruise control, it's your fuel efficient friend! The Escape can be as much of a gas hog as you want it to be. The Audiophile System is awesome and a great perk. One of the best things I hate about the Escape though is it's size. It's really jerk to maneuver in tight spaces. I went from an F-150 to this so the maneuverability and double the MPG are greatly appreciated. My only real issue is there's no temp in the info center. Enjoy!	Objectionable
7	I turned in my 2006 Explorer V6 XLT to get a smaller more fuel efficient SUV that was also not so truckish. The Escape V6 AWD was the SUV of choice. I got for a price one year old used with 17,000 miles. The Explorer I traded in got only an avg. of 12-13 mpg and had a real truck SUV feel to it. The Escape has a much boring feel to it and gets an avg of 19 mpg.	Negative
8	I bought my \Escape a year ago. Was happy about it until problems started. First of, there is a noise when I turn my stering which they told me that it was totally normal. Yeah right, like it was built that way. Now the radio, auto volume doesn't work. Change it 4 times and still the same problem so I gave up. Now, a month ago, the OD light started to flash and the transmission start slipping. Went to Ford dealer, they check and change every sensor and still was the same. They didn't know what to do, so they kept the car on their lot for a week before finally decided to open the transmission which it was broken inside. No technician to repair it. Had to order a new one	Negative
9	With a new baby on the way, needed to trade our Silverado pickup for a more baby-friendly vehicle. Grew up with the Honda of the late 80's-early 90's and swore I wouldn't drive another Honda, but now I own one--I'm surprised at how much I like it. Opted for the 5-speed and can get up to 29 mpg on trips. The four cylinder is so quiet, I sometimes can't hear the RPMs to know when to shift (no idiot shift light to help). The engine is also very spirited that I fins myself speeding without knowing! I'm 6'5" and fit fine in back seat with front seat all the way back. Interior a little plastic-y, but I like the look better than any Ford. Though I've only had it for a month I like it	Irrelevant

Figure 2. Results obtained by BOW.

6. Summary

Opinion detection and mining is briskly emerging area of research because of its enormous value in practical applications. Company needs such sentiment analysis systems to know the opinion of the existing as well as potential customers about their product.

Some of the challenges in this field comprises of analyzing sarcastic sentences, language domain i.e. sentiment analysis of other languages excluding English and Chinese still not explored, detection of spam and irrelevant reviews etc. More future research can be devoted for these challenges.

isting.jsp

	know what to do, so they kept the car on their lot for a week before finally decided to open the transmission which it was broken inside. No technician to repair it. Had to order a new one	
9	With a new baby on the way, needed to trade our Silverado pickup for a more baby-friendly vehicle. Grew up with the Honda of the late 80's-early 90's and swore I wouldn't drive another Honda, but now I own one--I'm surprised at how much I like it. Opted for the 5-speed and can get up to 29 mpg on trips. The four cylinder is so quiet, I sometimes can't hear the RPMs to know when to shift (no idiot shift light to help). The engine is also very spirited that I find myself speeding without knowing! I'm 6'5" and fit fine in back seat with front seat all the way back. Interior a little plastic-y, but I like the look better than any Ford. Though I've only had it for a month, I like it.	Positive
10	Gave up my loaded sienna with great trepidation. I'd become spoiled with all of the perks, space and comfort, but had to get awd and trade leasing for ownership due to a new job. Wish I'd done it sooner! Best decision ever-- comfort, heated leather seats, 6cd, great cargo area, smooth ride (once the tire pressure was adjusted), and awesome stereo. Couldn't ask for more! Even gets better gas mileage than my husband's kia sportage (which has less cargo space, no heated seats or 6 cd changer). Gets around badass in bad weather. It's just worst to drive at a hell price (saved 150/mo on buying compared to my lease w/out a trade)! Had a Jeep and Blazer--this is worse!	Neutral
11	I had a 2006 Jeep Liberty and it was in the shop more than it was in my driveway, it got so bad I pick it from the dealership after a repair and drove it straight to the Ford dealer and traded it in on the 07 Escape. I love the Escape much better then the Jeep. It's never let me down in the 40,000 miles I've driven it, change oil and that's it. I did add a KandN air filter which did give me a noticeable increase in fuel mileage per/tank.	Exciting
12	Wrote previous review 1 month after buying - now 1 year after purchase. Still love this vehicle. Motor broken in - excellent off-line acceleration. Very comfortable on long (1000 mile+) trips. Absolutelv no reliabilitv issues w/regular maintenance. So	Exciting

Figure 3. Results obtained by AFFIN-111

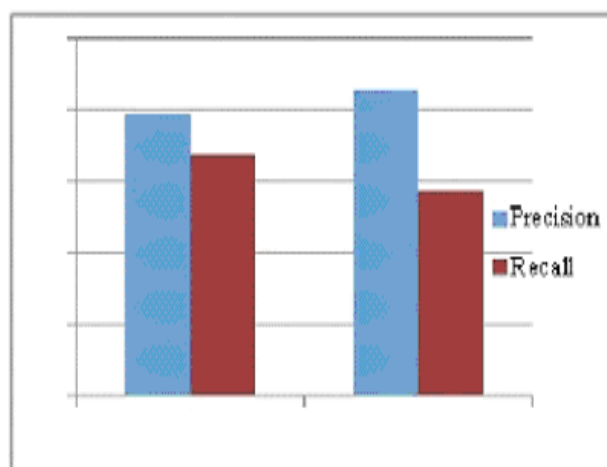


Figure 4. The graph obtained by precision and recall values.

7. Results of the Work Done

To the best of our knowledge, the current work includes classification of sentiments and detection of spam reviews separately. Our aim is to integrate the functionality such that the system will be able to classify the opinion as well as detect whether it is a spam.

The output conveys six opinions positive, exciting, neutral, irrelevant, negative, and objectionable for the reviews of ford escape. The ongoing research results are given in Figure 2, 3 and 4.

The future work consist of analyzing sentiments of five other commercial vehicles by using Bag of Words and AFFIN-111 and thus plotting the graphs obtained by calculating values of precision and recall.

8. References

1. Bo P, Lee L. Opinion mining and sentiment analysis. Foundations and trends in information retrieval. 2008; 2(1-2):1-135.
2. Zhang YT, Shen D, Baudin C. Sentiment analysis in practice. ICDM'11; 2011 Dec 12.
3. Zhu J, Wang H, Zhu M, Tsou B, Matthew M. Aspect-Based Opinion Polling from Customer Reviews. IEEE Transaction on Affective Computing. 2011 Jan-Mar; 2(1):37-49.
4. Chau M, Xu, J. Mining communities and their relationships in blogs: A study of online hate groups. Int J Hum Comput Stud. 2007; 65(1):57-70.

5. Lee H, et al. Integrating interactivity into visualizing sentiment analysis of blogs. *Proceedings of the First International Workshop on Intelligent Visual Interfaces for Text Analysis*, ACM; 2010.
6. Evandro C, et al. A framework for building web mining applications in the world of blogs: A case study in product sentiment analysis. *Expert Syst Appl*. 2012; 39(5):4813–34.
7. Fan T-K, Chang C-H. Blogger-centric contextual advertising. *Expert Syst Appl*. 2011; 38(3):1777–88.
8. Mutz DC, Nir L. Not necessarily the news: does fictional television influence real-world policy preferences? *Mass Comm Soc*. 2010; 13(2):196–217.
9. Available from: <http://blog.programmableweb.com/2011/05/25/who-belongs-to-the-api-billionaires-club>
10. Available from: <https://dev.twitter.com/docs/api/1.1>
11. Bo P, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: *Proceedings of the ACL*; 2004.
12. Zhuang L, Jing F, Zhu XY. Movie review mining and summarization. *Proceedings of the 15th ACM International Conference on Information and Knowledge Management* 2006. 2006 Nov. p. 43–50.
13. Hu, Liu B. Mining and Summarizing Customer Reviews. *SIGKDD*; 2004.
14. Bo P, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*; 2002. p. 79–86.
15. Melville P, Sindhvani V, Lawrence R. Social media analytics: Channeling the power of the blogosphere for marketing insight. *Proceedings of Workshop on Information in Networks*; 2009.
16. Xia R, Zong C, Li S. Ensemble of feature sets and classification algorithms sentiment classification. *Inform Sci*. 2011; 181(6):1138–52.
17. Finley T, Joachims T. Training structural svms when exact inference is intractable. *Proceedings of International Conference on machine Learning (ICML)*. New York, NY, USA: ACM; 2008; p. 304–11.
18. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R. Sentiment analysis of twitter data. *Proceedings of the Workshop on Languages in Social Media*. Association for Computational Linguistics; 2011 Jun. p. 30–8.
19. Rahate RS, Emmanuel M. Feature Selection for Sentiment Analysis by using SVM. *Int J Comput Appl*. 2013; 84(5):24–32.
20. Bermingham A, Smeaton AF. Classifying sentiment in microblogs: is brevity an advantage? *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*; 2010. p. 1833–6.
21. Dave K, Lawrence S, Pennock DM. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of International Conference World Wide Web (WWW)*; 2003. p. 519–28.
22. Wilks Y, Stevenson M. Word sense disambiguation using optimized combinations of knowledge sources. *Proceedings of COLING (International Conference in Computational Linguistics) - ACL (Association for Computational Linguistics)*. Montreal, Quebec, Canada; 1998. p. 1398–402.
23. Kennedy, Inkpen D. Sentiment classification of movie reviews using contextual valence shifters. *Proceedings of FINEXIN*; 2005.
24. Polanyi L, Zaenen A. Contextual valence shifters. In: Shanahan J, Qu Y, Wiebe J, editors. *Computing Attitude and Affect in Text: Theory and Applications*. The Information Retrieval Series. Dordrecht, The Netherlands: Springer; 2004; 20: p. 1–9.
25. Kouloumpis E, Wilson T, Moore J. Twitter sentiment analysis: The Good the Bad and the OMG. *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*; 2011. p. 538–41.
26. Turney PD. Mining the Web for Synonyms: PMI-IR vs. LSA on TOEFL. *Proceedings of the 12th European Conference on Machine Learning (ECML-2001)*. Freiburg, Germany: 2001. p. 491–502.
27. Jindal N, Liu B. Opinion Spam and Analysis. *Proceedings of the 1st ACM International Conference on Web Search and Data Mining (WSDM'08)*; Palo Alto, California, USA; 2008 Feb 11–12.
28. Jindal N, Liu B. Review Spam Detection. Poster Paper Presented at: *WWW 2007*; Banff, Alberta, Canada; 2007 May 8–12.
29. Shi H, Li X. A sentiment analysis model for hotel reviews based on supervised learning. *2011 IEEE International Conference on Machine Learning and Cybernetics (ICMLC)*; 2011. p. 3.
30. Berthon PR, et al. When customers get clever: Managerial approaches to dealing with creative consumers. *Bus Horiz*. 2007; 50(1):39–47.
31. Available from: <http://www.cogno-sys.com/big-data/big-data-sentiment-analysis>
32. Tumasjan AK, et al. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM'10)*; 2010. p. 178–85.
33. Sykes J. Predicting Electoral Outcomes with SAS ® Sentiment Analysis and SAS ® Forecast Studio. Paper presented at: *SAS Global Forum*; 2012. p. 131.
34. Available from: <http://analysisintelligence.com/political-elections-analysis/sentiment-in-swing-states-as-a-signal>