Sentiment Analysis on Product Reviews Using Machine Learning Techniques

Rajkumar S. Jagdale, Vishal S. Shirsat and Sachin N. Deshmukh

Abstract Sentiment Analysis and Opinion Mining is a most popular field to analyze and find out insights from text data from various sources like Facebook, Twitter, and Amazon, etc. It plays a vital role in enabling the businesses to work actively on improving the business strategy and gain an in-depth insight of the buyer's feedback about their product. It involves computational study of behavior of an individual in terms of his buying interest and then mining his opinions about a company's business entity. This entity can be visualized as an event, individual, blog post or product experience. In this paper, Dataset has taken from Amazon which contains reviews of Camera, Laptops, Mobile phones, tablets, TVs, video surveillance. After preprocessing we applied machine learning algorithms to classify reviews that are positive or negative. This paper concludes that, Machine Learning Techniques gives best results to classify the Products Reviews. Naïve Bayes got accuracy 98.17% and Support Vector machine got accuracy 93.54% for Camera Reviews.

Keywords Sentiment analysis · Natural language processing · Product reviews Machine learning · Support vector machine · Naïve Bayes

1 Introduction

Sentiment analysis invokes to the study of text analysis, natural language processing, computational linguistic to scientifically identify, extract and study subjective information from the textual data. Sentiment or opinion is the attitude of

R. S. Jagdale (✉) · V. S. Shirsat · S. N. Deshmukh
Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, 431004 Aurangabad, Maharashtra, India
e-mail: rajkumarjagdale@gmail.com

V. S. Shirsat
e-mail: vss.csit@gmail.com

S. N. Deshmukh
e-mail: sndeshmukh@hotmail.com

P. K. Mallick et al. (eds.), *Cognitive Informatics and Soft Computing*,
Advances in Intelligent Systems and Computing 768,
https://doi.org/10.1007/978-981-13-0617-4_61

customers comes from reviews, survey responses, online social media, healthcare media, etc. General meaning of sentiment analysis is to determine the insolence of a speaker, writer, or other subject with respect to particular topic or contextual polarity to a specific event, discussion, forum, interaction or any documents, etc. Essential task of Sentiment analysis is to determine polarity of given text at the feature, sentence, and document level. Due to increase in user of Internet every user is interested to put his opinion on the internet through different medium and this results opinioned data has generated on the internet. Sentiment analysis helps to analyze these opinioned data and extract some important insights which will help to other user to make decision. Social media data can be from different types like Product Reviews, Movie reviews, Reviews from airlines, Cricket Reviews, Hotel Reviews, employee interaction, Healthcare reviews, news and articles etc.

1.1 Data Sources

In Sentiment Analysis and Opinion Mining there are different Data Sources for generating huge amount of data on social media. Some are given below.

1.1.1 Blogs

It is website on which person can write their opinion on particular thing.

1.1.2 Datasets

Most of the researcher has uploaded different datasets online with free access. E.g. Movie reviews, Product reviews, Hotel reviews, etc.

1.1.3 Review Sites

Most of the customers are putting their opinion on E-commerce websites where they brought products. Like Amazon, CNET, epinion, zdnet, consumerreview, IMDB, etc.

1.1.4 Micro-blogging

It is popular service for sending text message shortly. Like Twitter, Tumblr, Dipity, etc. (Fig. 1).

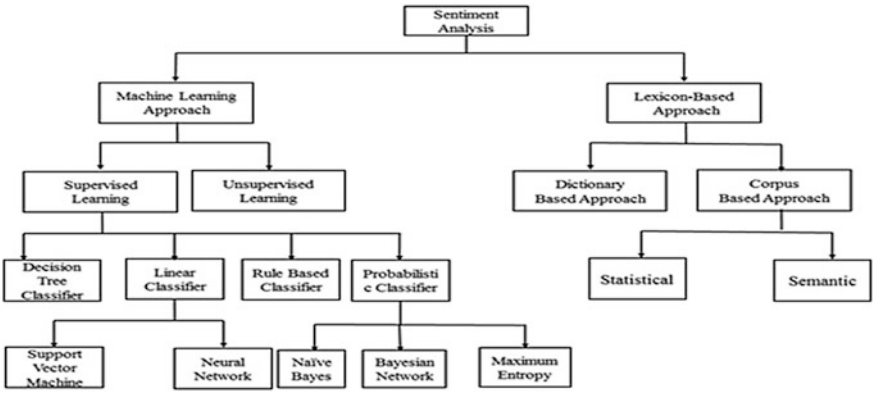

Fig. 1 Different sentiment analysis technique

2 Related Work

Sentiment Analysis is top research field under natural language processing (NLP), containing process of detecting and extracting sentiment/opinion from the text and classifying their sentiment. Sentiment analysis studies people's opinion, appraisal, emotions, and attitude towards individual, organization, products, movies, issues, events, etc.

2.1 Document Level

As per name, it analyzes the documents. In this level whole document has been analyzed and classify that whole document is [1, 2] expressing positive or negative view. In this method only one product reviews has processed and find out the opinion about same product. Opinions are expressed on only single entity. If document has multiple product reviews then this level will not work because it is not relevant to that documents which contains multiple type product reviews.

2.2 Sentence Level

In this level task goes to the sentences and analyzes then determines whether the given sentence has positive, negative, and neutral opinion. This level somehow similar to Subjectivity Classification [3] which separates the objective sentences and subjective sentences. Factual information and subjective information containing sentences called Objective sentences and Subjective sentences respectively.

2.3 Aspect Level

Earlier, Aspect level sentiment analysis is called Feature level sentiment analysis [3] contains feature-based opinion mining and summarization. To find out what exactly people like or did not like this level is very important. It is finer-grained sentiment analysis level. Aspect level directly looks at the opinion itself instead of looking to document or sentences. Output of this level will be the entity, aspect of that entity, opinion of aspect, opinion holder and time. For example, Samsung j7 has best camera quality. Now in this level, camera is aspect of entity Samsung j7 which expresses the positive opinion. In movie, particular scene, actor, acting, actions, etc., are some key points under aspect level sentiment analysis.

They compared various for Sentiment Analysis [4] by examining different methodologies and discussed Machine Learning methods like SVM, NB, and ME. They also discussed N-gram for Sentiment Analysis. Two steps method (Aspect Classification and polarity detection) has been used [5] and discussed lexical approach. Customer reviews are used as a dataset in their proposed method and in experimental result. SVM has been used and achieved 78% accuracy. Datasets consist of tweets [6] annotated for sentiment on a 2-point, 3-point, and 5-point scales. This is twitter task from SemEval 2017 and has done sentiment analysis on a two-point and on a five-point ordinal scale. They used new language, Arabic, for all subtasks. Polarity Categorization problem has been solved [7] for Sentiment Analysis. Online product reviews from Amazon are used for experimental purpose. Sentence level categorization and review-level categorization are implemented in this paper.

Hybrid approach [8] has been used, i.e., combination of Machine Learning and Lexicon based approach. In this paper, different techniques and tools have been discussed with different aspects. pSenti, combination of Lexicon-Based- and Learning-Based methods for Twitter Sentiment Analysis, SAIL, NILC_USP, combination of Lexicon-based- and Learning-based approaches for improved performance and convenience in sentiment classification, A Hybrid approach for sentiment classification of Egyptian Dialect Tweets, Sentiment Analysis: A Review and Comparative Analysis of Web Services, Alchemy API, Building Large-Scale Twitter-Specific Sentiment Lexicon: A Regression Learning Approach, Sentiment Analysis on Twitter, Sentiment Analysis using Sentiment Features, Improving Twitter Sentiment Analysis with Topic-Based Mixture Modeling and Semi-Supervised Training and MSA-COSRs are used as Hybrid tools and techniques. They surveyed on Opinion mining [9] with respect to their different levels, tools used, architecture, techniques applied, comparative study of techniques and challenges.

Sentence level sentiment analysis [10] has been done and taken live tweets from twitter using R tool. Also discussed different lexicons Like SentiWordNet, WordNet-Affect, MPQA, etc. Tweets of different events has been collected like #Budget2016, #RailBudget2016, #Freedom251, #MakeInIndia, #Oscars2016, #startup, #InternationalWomensDay, #AsiaCupT20Final, #IndvsPak and

#ProKabaddi. Each event has 10,000 tweets and classified as positive and negative tweets which gives information about people's opinion about that events. Machine learning approaches [11] used and also discussed spammed reviews and unauthenticated users. Amazon review dataset has been used in this paper. Review contains Reviewer ID, Product ID, Review Text, Rating and time of the review. They identified public perception of their product over time and discovered important areas where their products can be more improved.

3 Proposed Method

In the proposed method following preprocessing task has been completed to classify sentiment analysis from reviews.

3.1 Collection of Dataset

The dataset is collected from Amazon and it is in json format. Each json file contains number of reviews. Dataset has reviews of Camera, Laptops, Mobile phones, tablets, TVs, video surveillance.

3.2 Preprocessing

In preprocessing tokenization, stop word removal, stemming, punctuation marks removal, etc., has done. It has converted in bag of words. Preprocessing is important in sentiment analysis and opinion mining.

3.3 Score Generation

In this step, every sentence has analyzed and calculated sentiment score. To calculate sentiment score dataset has compared with opinion lexicons i.e. 2006 positive words and 4783 negative words and calculated sentiment score for every sentence.

3.4 Sentiment Classification

Using score and different features different machine learning algorithms has applied and different accuracy measurements calculated. Proposed method uses the

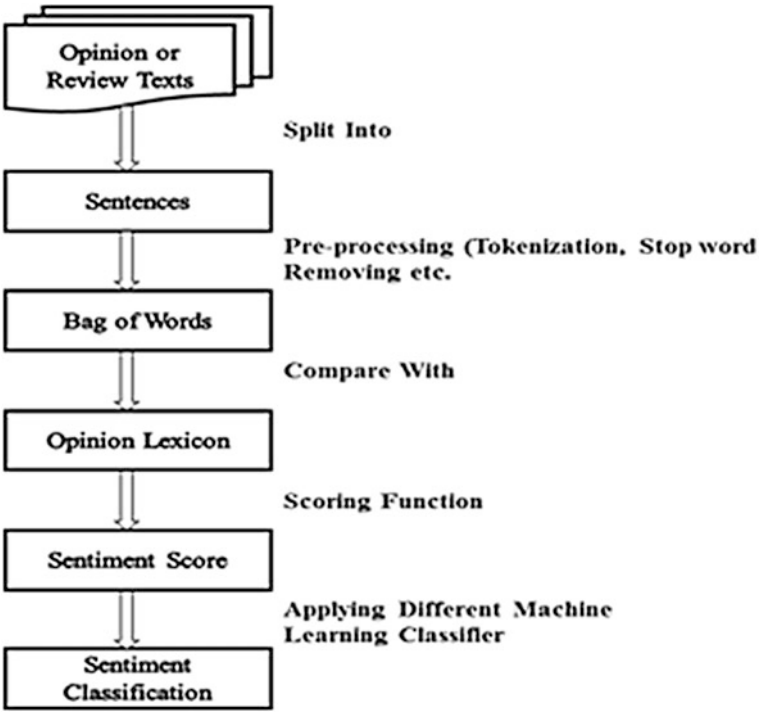

Fig. 2 Proposed methodology of sentiment analysis

following tasks to classify the sentiment analysis using machine learning techniques (Fig. 2).

4 Experimental Results

4.1 Dataset Description

Dataset contains reviews of Camera, Laptops, Mobile phones, tablets, TVs, video surveillance collected from Amazon. It is in the form of json files and each json file contains number of reviews (Fig. 3; Table 1).

4.2 Classification Results

See (Fig. 4).

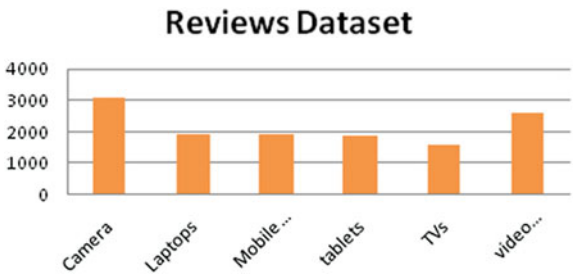

Fig. 3 Graphical representation of number of reviews

Table 1 Dataset and reviews count

Dataset name	Number of reviews
Camera	3106
Laptops	1946
Mobile phones	1918
Tablets	1894
TVs	1596
Video surveillance	2597

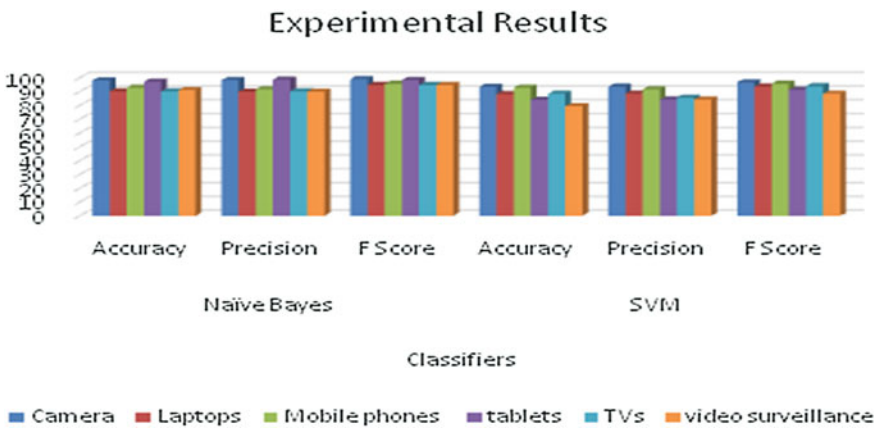

Fig. 4 Graphical representation of experimental results

Table 2 Evaluation parameters for classifiers of datasets

Dataset	Classifiers					
	Naïve Bayes			SVM		
	Accuracy	Precision	F score	Accuracy	Precision	F score
Camera	98.17	98.30	99.03	93.54	93.58	96.66
Laptops	90.22	90.01	94.74	88.16	88.52	93.71
Mobile phones	92.85	91.64	95.64	92.85	91.64	95.64
Tablets	97.17	98.73	98.31	84.12	84.31	91.37
TVs	90.16	90.17	94.72	88.49	85.56	93.89
Video surveillance	91.13	89.95	94.71	79.43	84.25	88.53

5 Conclusion and Future Work

Nowadays, Research on Sentiment Analysis and Opinion Mining is very important. Most of industries are creating different types of data and they need to analyze that data to make decisions which are benefited to industry. Social media is also generating huge amount of data and that need to analyze and find insights from that data. In this paper product reviews Dataset has taken from Amazon website containing six product reviews of Camera, Laptops, Mobilephones, Tablets, TVs, and Video Surveillance. In this proposed methodology dictionary-based approach under lexicon-based Approach has been used with machine learning techniques. Sentiment Analysis has been done on every product reviews and then classified using machine learning algorithms, i.e., NB and SVM. Table 2 shows the accuracy measurements for NB and SVM classifier for the dataset. Naïve Bayes classifier got 98.17% accuracy for Camera reviews and Support Vector Machine got 93.54% accuracy for Camera reviews. For future work, on same dataset, Aspect level sentiment analysis can improve results and from that method, we can get what exactly people liked or disliked. For example, Camera's quality, megapixel, picture size, structure, lens, picture quality, etc. Aspect level is fine-grained approach and gives good result in Sentiment Analysis.

Acknowledgements The author would like to thank to Data Analytics Research Lab, Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad for providing infrastructure to carry my research work. The author acknowledges the Department of Science and Technology (DST), New Delhi, India for granting financial assistance in the form of DST INSPIRE FELLOWSHIP (JRF) during this research work.

Bibliography

1. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10. Association for Computational Linguistics, 2002
2. Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002
3. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2004
4. Devika, M.D., Sunitha, C[a]., Ganesh, A.: Sentiment analysis: a comparative study on different approaches. Proc. Comput. Sci. **87**, 44–49 (2016)
5. Bhadane, C., Dalal, H., Doshi, H.: Sentiment analysis: measuring opinions. Proc. Comput. Sci. **45**, 808–814 (2015)
6. Rosenthal, S., Farra, N., Nakov, P.: SemEval-2017 task 4: sentiment analysis in Twitter. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (2017)
7. Fang, X., Zhan, J.: Sentiment analysis using product review data. J. Big Data **2**(1), 5 (2015)
8. Ahmad, M., et al.: Hybrid tools and techniques for sentiment analysis: a review. Int. J. Multidiscip. Sci. Eng. **8**(3) (2017)
9. Sharma, S., Tiwari, R., Prasad, R.: Opinion mining and sentiment analysis on customer review documents—a survey. IJARCCE **6**(2), 156–159 (2017)
10. Jagdale, R.S., Shirsat, V.S., Deshmukh, S.N.: Sentiment analysis of events from Twitter using open source tool. IJCSMC **5**(4), 475–485 (2016)
11. Kamalapurkar, D., Bagwe, N., Harikrishnan, R., Shahane, S., Gahirwal, M.: Sentiment analysis of product reviews. Int. J. Eng. Sci. Res. Technol. **6**(1), 456–460 (2017)