

A Review on sentiment analysis techniques and their performance

Kanwaldip Kaur¹, Research Scholars (Deptt. of Computer Science)
Dr. Rajan Manro ², (Deptt. of Computer Science),Khanna(Punjab)

Abstract: In today's time various people give their reviews about their sentiments after shopping or even various social issues. Using large processing engines these reviews can be interpreted to generate the analysis for the positive and negative polarity. Various researchers have focused on different types of techniques and tools for the analysis purpose. These analyses will be useful for the purpose of decision making for different types of organizations. Researches are based on studying the reviews on tourism, shopping, Govt. decisions etc.. With the help of lexical approach tokenize the reviews. Identify the positive, negatives and neutral reviews by comparing them with established Ontology. This mined information will provide a better view regarding the total reviews. Rather than studying all the individual reviews, it will be much better to have a collective analysis by an automated tool.

Key Terms: Ontology, Lexical, Reviews.

I. INTRODUCTION

Web produces an enormous amount of data. This data grows every day. A large chunk belongs to people's reviews for different products they had purchased previously through E-commerce or even what they feel about the government decisions. These reviews are regarding their experience of product purchases or opinion about the decision of any individual. These reviews sometimes can be false and influenced. Various people before purchase anything on E-commerce want to know the reviews of people. Now the problem is it is very difficult to see this enormous data and understand the positivity and negativity of the different reviews. So there is a hard requirement of a certain automated tool, which can process this large amount of data and extract useful trends from all these reviews, which will be much easier to understand the percentage of positive and negative reviews. In the past decade or so E-Commerce industry on internet has

risen many folds. Online shopping has grown tremendously. Doing so will provide time-efficient solutions. There are high chances that people get cost-effective products, cost-effective foods etc. Through online procedure shops. Later on they give reviews regarding the product and services they get once they experience in real. All the shopping websites also emphasize on people's post-experience. Because a study shows that people's purchase is influenced by these reviews. Once a certain automated tool will be available, it will provide a better way to search the relevant data and analysis with less time and cost. These automated tools will provide data graphically. A simple number on a rating system is not providing enough information, but neither a long review in which users express opinions about more than shopping features[1]. There are plenty of problems that exist in such a reviewing system like.

1. Long reviews sometimes become senseless.
2. Rating-based reviews are less understandable.
3. Reviews sometimes are misleading.
4. Reviews sometimes are biased.
5. Reviews sometimes provide a duality of views.
5. People sometimes has a language problem in both reading and writing.

A system that will analyze the reviews and also extract the useful analysis from the reviews data. This analysis will be useful for consumers who usually put the shopping based on people's reviews. But it will be difficult to read each individual review. So a collective analysis will help in understanding the things and put the general point of view.

I.1 Analysis Module

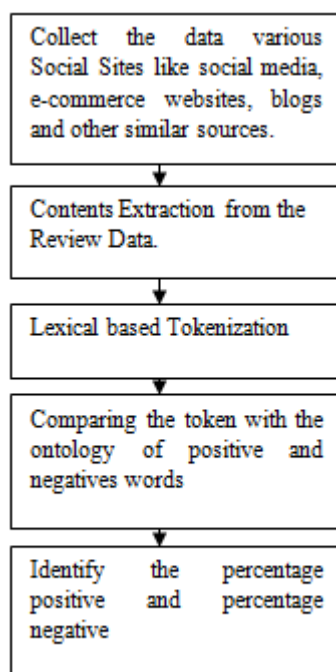


Figure 1. Review analysis Flow [1]

1.2 Methodology

The process of sentiment analysis includes various sequential steps. These steps are like

- a. data acquisition,
- b. data pre-processing and normalization
- c. feature extraction and representation,
- d. Labeling

1.3 Tools

Which having sentiment analysis various tools are used

1. Hadoop(Hive).
2. Matlab.

II. RELATED WORK

SHAHID SHAYAA(2018) et. al: this paper focused on sentiment analysis for the big data. opinion mining and sentiment analysis for the Big data categories the people sentiment into different classes. Each class represents the people mood for the current aspect under consideration. This paper is focused on technical and non technical aspect of the system for the opinion mining.

Swati Redhu(2018) et. al: This paper focuses on the process of opinion mining and sentiment analysis for the Big data. The data collected is from different sources like social media, e-commerce, blogs, social media. This paper focuses on data

acquisition, data pre-processing, Normalization and feature extraction and finally representation. This paper focuses on the different methods and techniques for the sentiment analysis and opinion mining.

DAVID ZIMBRA (2018) et al: This paper focuses on the process of classification for the sentiment analysis for different classes of the twitter data into different classes. The success rate of different classification is around 70%. 28 different commercial and academic system classifies different datasets into different classes. Paper focuses on the different trends and techniques for improving the results.

M.Trupthi(2017)et. al: This paper is focused on sentiment analysis of large number of reviews or tweet data on various social media sites. Large amount of data is being processed by the processing engine. This processing engine can be Hadoop etc. This can process the large amount of data to extract the useful patterns. Sentiment analysis in current paper is based on features.

Prerna Mishra(2016)et. al: This paper focus on sentiment analysis for the tweeter data. Paper has collected a data of the persons opinion on Modi ji's digital india campaign. The data collected is from various social media tools. First they have collected the sentiment of the persons. In second step classifying the data based on polarity of these sentences w.r.t. Positive, Negative, or neutral. From the analysis it is clear that the positive polarity is 50%, negative polarity as 20% and neutral polarity is 30%.

Cristian Bucur(2015) et al: This paper has proposed a robust system for analysis the review data taken from various reviewing sites. These reviews sites are specifically related to tourism. The researcher is of point of view that the analysis and representation in the advantage of both customer and service provider. Because customer can choose best at best price. Service provider can better there service. Because it is always learning to read direct reviews about your services. You can always fine tune the services if certain shortfall exists. But simultaneously he is of point of view that review should not be biased. Because it leads to wrong analysis.

Aliaksei Severyn(2015) et al: According to this researcher large amount of data is produced every day on the internet. This data is regarding reviews of the data. There requires robust model which can have transformation procedure for these reviews data. So that data can be analysis and be converted to well under stand able way. This is the data related to you-tube videos data. These reviews are the people personal views on the reviews

Kamal MAROUCHE(2015) et al: Today's environment is related to competition. Organization compete amongst themselves for attaching more and more customer. Because they have online stores. People do visit there stores repetitively. So companies wants to attach their customers to them by taking there online reviews. They carefully analysis the reviews data and put it in right perspective. Once the robust analysis is complete they put corrective actions to fine tune their services. So that they can attract new customer. The

robust tool for analysis purpose will helps in having system more understandable and automatic

Montoyo, Martínez-Barco, & Balahur(2012). Majority of researchers are based on opinion mining for standardized text. This standardized data includes shopping reviews and news data Amazon data, and various type of other data. But the problem is the words and sentences used into these reviews are not standardized reviews. It will be very difficult to process this language containing informal words. But one thing in common is that all the reviews data is in English.

III Methodology

- data acquisition
- data pre-processing
- normalization
- feature extraction
- Representation,
- labeling

IV. Comparison Table

Author Name	Name of Paper	Technique	Constraints
Kamal AMAROUCHE et. al(2015):	Product Opinion Mining for Competitive Intelligence	Classification based technique. Classify on the basis of Machine learning, Lexicon Approach.	The approach can be further enhanced to increase the percentage of true positive and false negatives.
Cristian Bucur(2015)	Using Opinion Mining Techniques in Tourism	It is based on classification and aspect based approach	Further there requires to improve the performance by using ontology.
J. Islam(2016) et. al.	An Effective Approach of Intrinsic and Extrinsic Domain Relevance Technique for Feature Extraction in Opinion Mining	This paper has applied the approach of NLP that is natural language processing. It also uses Intrinsic and Extrinsic Domain Relevance.	This research can be further enhanced to consider two thresholds for better analysis.
S. S. Sohail(2015) et. al	A model for enhancement in wearable technology based on opinion mining technique	Any user which provide feedback will be useful aspects for many processing and analysis techniques.	Till now large amount of information domains are untouched. Which can be further enhanced and processed in further researches.

S. Ahmed	A Novel Approach for Sentimental Analysis and Opinion Mining based on SentiWordNet using Web Data	Used machine learning algorithms like Naïve Bayes, SVM and Multilayer Perception (MLP)	Work can be further enhanced by selecting the methods based on threshold like info gain and chi square tests.
Dhanalakshmi V.(2016) et al.	Opinion mining from student feedback data using supervised learning algorithms	comparative performance study of the algorithms like SVM, Naïve Bayes, K Nearest Neighbor and Neural Network classifier.	The work can be further enhanced by considering the more suitable techniques for categorization. These analysis tools will be based on SVM and Naïve Bayes..

V. CONCLUSION

Studying different papers it is clear that reviews on different products by customer in different ways influence the purchase of the customers. Different ways of reviews can be star rating, Reviews in sentences, or even yes or no. Different researchers has studied these reviews in different ways. Also they have studied the effect of reviews on the future purchase of other customers. Various researchers has emphasis on processing the reviews data. So that any user rather than seeing the enormous data can see the processed data and make their mind more easily and effectively. Authors has emphasis the point that certain automated tool is required which can process this enormous data and extract the use full data. This useful data will be presented in useful and understandable way. Simultaneous to the positives there are various negatives. These negatives are due to biased reviews and wrongly presented reviews. So large amount of research further can be done, which can helps in removing these constraints and provides further reviews in useful way.

VI. FUTURE WORK

Study of various research techniques has given the procedure of sentiment analysis for given set of reviews given by the people on different subjects. These techniques has different success rate for the identification of true positive, false negative etc.. Further the technique can be enhanced for the sentiment analysis for the review data based on lexical

analysis and generating the ontology of large number of the positive and negative sentences. This will enhanced the success rate of the generating the true positive and true negatives.

REFERENCES

- Shahid Shayaa1 , Noor Ismawati ,Jaafar , Shamshul Bahri , Ainin Sulaiman, Phoong Seuk Wai, Yeong Wai Chung , Arsalan Zahid Piprani, And Mohammed Ali Al-Garadi, "Sentiment Analysis Of Big Data: Methods, Applications, And Open Challenges",Ieee,Vol. 6 2018.
- Swati Redhu, Sangeet Srivastava1, Barkha Bansal , Gaurav Gupta," Sentiment Analysis Using Text Mining: A Review", International Journal On Data Science And Technology,Vol. 4(2),Pp: 49-53, 2018.
- David Zimbra, Ahmed Abbasi, Daniel Zeng, Hsinchun Chen," The State-Of-The-Art In Twitter Sentiment Analysis: A Review And Benchmark Evaluation", Acm,Vol. 9, 2018.
- M.Trupthi, Suresh Pabboju, G.Narasimha," SENTIMENT ANALYSIS ON TWITTER USING STREAMING API", IEEE,2017.
- Perna Mishra, Dr. Ranjana Rajnish, Dr.Pankaj Kumar," Sentiment Analysis of Twitter Data:Case Study on Digital India",IEEE,2016.
- Cristian Bucur," Using Opinion Mining Techniques in Tourism", Procedia Economics and Finance 23 (2015) 1666 – 1673.

- Aliaksei Severyn a, Alessandro Moschitti c,a,†, Olga Uryupina a, Barbara Plank b, Katja Filippova,” Multi-lingual opinion mining on YouTube”, Information Processing and Management xxx (2015) xxx–xxx.
- Kamal AMAROUCHE, Houda BENBRAHIM, Ismail KASSOU,” Product Opinion Mining for Competitive Intelligence”, Procedia Computer Science 73 (2015) 358 – 365
- Bartes, F. Action Plan Basis of Competitive Intelligence Activities. Economics and Management, 1, (2011) 664-669.
- Safarina, H., Akbari, Z. & Abbasi, A. Review of Competitive Intelligence & Competitive Advantage in the Industrial Estates Companies in the Kerman City. International Business and Management, 2, (2011) 47-61.
- Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. Computational Linguistics, 34(4), 555–596.
- Baccianella, S, Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In LREC.
- Baldwin, T., Cook, P., Lui, M., MacKinlay, A., & Wang, L. (2013). How noisy social media text, how diffrent social media sources? In IJCNLP.
- Basile, V., & Nissim, M. (2013). Sentiment analysis on Italian tweets. In 4th Workshop on computational approaches to subjectivity, sentiment and social media analysis.
- Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In ACL.
- Bosco, C., & Lombardo, V. (2006). Comparing linguistic information in treebank annotations. In LREC.
- Bosco, C., Patti, V., & Bolioli, A. (2013). Developing corpora for sentiment analysis: The case of irony and senti-TUT. IEEE Intelligent Systems, 28(2), 55–63.
- Cambria, E., & Hussain, A. (2012). Sentic computing: Techniques, tools, and applications. Springer.
- Carvalho, P., Sarmiento, L., Silva, M.J., & de Oliveira, E. (2009). Clues for detecting irony in user-generated contents: oh..!! it’s so easy;-). In 1st international CIKM workshop on Topic-sentiment analysis for mass opinion.
- Daumé III, H. (2007). Frustratingly easy domain adaptation. In ACL.
- Gimpel, K., Schneider, N., O’Connor, B., Das, D., Mills, D., Eisenstein, J., et al. (2011). Part-of-speech tagging for Twitter: Annotation, features, and experiments. In ACL.