

Survey on Sentiment Analysis: A Comparative Study

Himadri Tanaya
Chidananda
M.Tech, School of Computer
Engineering KIIT University,
Bhubaneswar, India

Santwana Sagnika
Professor, School of Computer
Engineering KIIT University,
Bhubaneswar, India

Laxman Sahoo
Professor, School of Computer
Engineering KIIT University,
Bhubaneswar, India

ABSTRACT

Sentiment Analysis (SA) is one of the greatest broadly planned applications of Natural Language Processing (NLP) and Machine Learning (ML). This field has grown enormously with the advent of the Web 2.0. The Internet has as long as a platform for people to express their opinions, emotions and feelings towards products, persons, and life in general. Accordingly, the Internet is nowadays a massive resource of opinion amusing written data. A vital job of sentiment analysis is sentiment classification, which intentions to automatically classify opinionated text as being negative, positive, or neutral. This paper provides a comparative study on sentimental analysis and its applications mostly for recommendation system. Recommender systems have grown to be a serious research area after the emergence of the first paper on collaborative filtering in the Nineties.

Keywords

Big Data, Recommendation system, Sentiment Analysis, Literature Review, Collaborative filtering, Content-based Filtering

1. INTRODUCTION

Data is the backbone of the 21st century, and analytics is the combustion engine. Data is everywhere, in every industry in the form of videos, images, numbers, and text. There is no restriction of the information as it shows everywhere throughout the universe. As data continues to grow, so does the need to sort it out. Every second, there are around 8,23 tweets on Twitter, every minute, nearly 510 comments are posted, 294,000 statuses are updated, and 136,000 photos are uploaded on Facebook, every hour, Walmart, a global discount departmental store chain, handles more than 1 million customer transactions. Collecting such large amount of data would just be a waste of time, storage space and effort if it cannot be put to any logical use. Organizations, independent entities, government, political parties and police, among others, are finally investing time and money in unlocking the power of data. They analyze the data to understand and interpret market trends, study customer behavior, and take financial and logistical decisions. The need to sort, analyze, organize and offer this critical data in a systematic manner leads to the rise of the much-discussed term, Big Data.

The term Big Data is now widely used particularly in the IT industry, where it has generated various job opportunities. The first organizations to embrace it were online and startup firms. Firms like LinkedIn, Google, eBay and Facebook were built around big data in the beginning. 'Big Data' is comparable to 'small data', but bigger in size. According to IBM, Big data is being generated by nearly everything around us at all times at an alarming velocity, volume, and variety. To extract meaningful value from Big Data, we need optimal processing

power and skills. As they live in the twenty-first century they exist with a new concept of data which is known as big data. There are three types of data we need to consider in Big Data: structured, unstructured, and semi-structured.

A. Structured Data

Structured data can be defined as the data that has a defined repeating pattern. This pattern makes it easier for any program to sort, read and process the data. Processing Structured Data is much easier and faster than processing data without any specific repeating patterns.

Example: MySQL, Oracle

B. Unstructured Data

Unstructured Data is a set of data that OR might not have any logical or repeating patterns. It consists typically of metadata, i.e. the additional information related to data, data in different formats such as e-mails, text, audio, video, or images.

Example: text, audio, video, opinion, Facebook

C. Semi- Structured Data

Semi-Structured Data, also known as having a schema-less or self-describing structure, refers to a form of structured data that contains tags or markup elements for different elements to separate elements and generate hierarchies of records and fields in the given data. Such type of data does not follow the proper structure of data models as in the relational database.

Example: CSV but XML and JSON documents are semi-structured documents, Sensor data, NoSQL databases are considered as semi structure.

1.1 Recommendation systems

Recommendation systems (RS) are a subclass of information filtering system that needs to predict the "rating" or "preference" that a user would give to an item. Recommender Systems are software tools and techniques providing suggestions for items to be of use to a user. Depending upon the product profile and the user profile, which are formed using various techniques and algorithms, suggestions are made. It gives the ability to recognize a user's taste and find original, required content for them consequently created on the pattern of their preferences and rating of various things [1]

Types of Recommendation systems

Over the years, recommender systems have been studied broadly and are divided basically into two types as follows: collaborative filtering (CF), content-based filtering. One more approach has turned out to be more prevalent where both the filtering methods can be useful in various stages of the recommendation system, known as hybrid filtering method.

1.1.1 Collaborative Filtering

Collaborative filtering methods construct a model from a customer's past behavior (items before pick up or particular and/or numerical ratings specified to those items) and in addition comparable choices by different clients, and use that model to forecast items (or ratings for items) that the customer might be interested in. e.g. Foursquare. [2], [3]

1.1.2 Content-based filtering

Content-based filtering is another methodology for recommender system. It is a technique where individual user profiles are taken into account. These techniques depend on account of the item and customer preferences. The keywords that are used to define the item show their features that used for producing recommendations. These algorithms also try to prescribe things that the user was interested in earlier. As such, different candidate items are compared with items previously rated by the user and the best-matching items are recommended. [2], [3] e.g. Flipkart.

1.1.3 Hybrid Filtering

Hybrid recommender systems combine two or more recommendation techniques to increase the efficiency of execution with fewer of the drawbacks than any individual one. Most commonly, collaborative filtering is combined with some other technique in an attempt to avoid the ramp-up problem. [4]

1.2 Sentiment analysis

Sentiment analysis also known as opinion mining is the field of study that uses people's thought, ratings, belief, feelings towards products, services, and individuals. It mainly focuses on positive and negative emotions of the user. They are very important in the current scenario because, a large number of user opinionated texts are available on the web now. This is a difficult problem solve because natural language is highly unstructured in nature. [5, 16]

Different Levels In Sentiment Analysis

Different levels of sentiment analysis are of three types [17]

1. Document Level: The whole file contains group opinion. The file verifies whether it conveys the positive or negative sentiment.
2. Sentence Level: Verifies that whether a sentence conveys positive, negative or neutral meaning.
3. Entity and Aspect Level: Verifies the emotions which are present at each level.

2. LITERATURE REVIEW

Recommender systems are developing as an effective tool in both industry and academia. Sentiment analysis aims to identify and extract opinions, moods, and attitudes of Individuals and communities.

In [6] author proposed a personalized recommendation systems. They design a Recommendation system via Learning automata and sentiment analysis. Here LASA framework is used which recommends the places adjacent the present location of the users. Average penalty is calculated using probability vector. It has strong merits like 1. It supports to get personalized recommendations built on the previous knowledge of the users. 2. LASA recovers the effectiveness of the recommender system. Demerits are 1. Secure aspects of recommender systems.

In [7] authors have put forth a KASR method for the personalized recommendation. In this paper user based

collaborative filtering algorithm is used. To make the technique more effective and scalable it is executed on Hadoop. Jaccard coefficient and Cosine similarity measure are used for evaluation. They show that the proposed recommendation method is better than the existing traditional methods. The major advantages are 1. Scalable 2. More efficient than traditional methods. There are also some demerits associated with this like 1. Jaccard Coefficient method is not so accurate. 2. User's positive and negative reviews are not separated. Sentiments in the text are not considered for calculation.

In [8] authors propose an active web service recommendation. Web utilization history and quality of service are the primary criteria for recommendation. Using this approach top k services are generated for users. Utilization history number is utilized for ranking. Merits are 1. Higher recall ratio and precision. 2. Show the power of the relationship among users. Demerits are Passive users reviews about the site is not considered. Usage history count is only used for ranking.

In [9] authors propose quality of service ranking prediction for cloud services. Rating based Approaches and ranking based approaches are studied in this paper. Merits are 1. The users can obtain Quality of Service ranking prediction as well as detailed Quality of Service value prediction. Demerits are 1. Applications in other field need further verification.

In [10] author proposed a such a social framework, which abstracts user's reviews, comments of eating place and points of attention such as events and places, to personalize and rank suggestions built on user first choice. In this Machine learning as well as sentiment analysis methods are used. Merits are 1. Unwanted data are removed 2. It enables faster services, better results. Demerits are 1. Accuracy can be increased by applying both contexts based and preference based searching.

In [11] authors propose a Bayesian inference based recommendation in online social networks. In this content, ratings are shared with friends. Conditional probability is used for calculating rating similarity. Based on similarity score ranking is done. They show that the proposed Bayesian inference-based recommendation is better than the existing trust-based recommendation. Merits are 1. Greater accuracy through friends' Recommendation. 2. Solve the difficult of a large size of an element in collaborative Filtering recommendation. Demerits are 1. There are a Cold start and rating sparseness issue.

In [12] author proposes a technique to acquire user's feedbacks in the scenario where the most user fail to give feedbacks, by automatic extraction of data about the specific product from twitter. They compare various text classifier technique like naïve Bayes and found SVM to be 82% accurate, so they used SVM as text Classifier. They used a third API twitter 4j for data extraction. Merits are 1. More accurate; 2. The rating can be done with the lack of direct information; Demerits are 1. Application in other field needs future investigation.

In [13] authors propose recommender scheme on behalf of game videos, transmitted over the network and communicate, in the context of across-the-board events, which has stayed tried on behalf of Olympic Games. The recommendation is depends proceeding audio visual consumption also not on the quantity of users, running merely on the client side. Merits are 1. This keeps away from the computation, concurrence, and security issue of central server methodologies in consequences with a huge number of users, example the

Olympic Sports event. Entire video has to recommend. Demerits are 1. Particular video fragment can't be recommended using this approach.

In [14] authors proposed a probabilistic personalized travel recommendation model. For mining demographics for travel landmarks and paths individuals attributes and photographs are utilized which are effective, and along these benefiting personalized travel recommendation services. Merits are 1. It gives more satisfactory Results.2. It improves group recommendations by mining travel preferences. Demerits are 1. The more competitive recommendation models need to be investigated.

In [15] authors proposed an exact investigation of the effectiveness of machine learning methods in categorizing instant memos by semantic significance. They propose different approaches in extracting text features such as bag-of-words model, handling negation, controlling to adverbs and adjectives hurdling word occurrences by a threshold, WordNet substitute's knowledge. The performance is estimated on an accuracy of four machine learning techniques are Decision Trees, K-Means clustering, Maximum - Entropy, Naive Bayes. Merits are 1. It evaluates the fitness of different feature selection and learning algorithms on the classification of comments according to their subjectivity and their polarity.

Table 1.Comparison Table

Paper name	Techniques	Merits	Demerits
Learning Automata Based Sentiment Analysis for Recommender System on Cloud	Here LASA framework is used.	1. It supports to get personalized recommendations built on the previous knowledge of the users. 2 .LASA recovers the effectiveness of the recommender system.	1. Secure aspects of recommender systems.
KASR: A Keyword-Aware Service Recommendation Method on Map Reduce for Big Data Applications.	User-based collaborative filtering algorithm is used. Jaccard coefficient and Cosine similarity measure is used for evaluation.	1. Scalable 2.More efficient than traditional methods	1. Jaccard Coefficient method is not so accurate. User's positive and negative reviews are not differentiated. Sentiments in the text is not considered for calculation
AWSR: Active Web Service Recommendation Based on Usage History	Web usage history and QoS are the main criteria for	1. Higher recall ratio and precision. 2. Show the	1. Passive user's reviews about the website are

	recommenda tion	power of the relationship among users.	not considered. Usage history count is only used for ranking.
Quality of Service Ranking Prediction for Cloud Services	Rating based Approaches and ranking based approaches are studied in this paper.	1. The users can obtain Quality of Service ranking prediction as well as detailed Quality of Service value prediction	1. Applications in other field need further verification
A Personalized Recommender System using Machine Learning based Sentiment Analysis over Social Data	In this paper machine learning and Sentiment Analysis, based methods are recycled to Observed accuracy results.	1.Unwanted data are removed 2. It enables faster services, better results	1. Accuracy can be increased by applying both context based and preference based searching
Bayesian-inference based recommendation in online social networks	In this content, ratings are shared with friends. Conditional probability is used for calculating rating similarity.	1. Greater accuracy through friends' Recommendation. 2. Solve the difficult of large size of element in collaborative Filtering recommendation	1. There are a Cold start and rating sparseness problem.
Product Rating Using Sentiment Analysis	In this paper by automatic extraction of data about the specific product from twitter. They compare various text classifier technique like naïve Bayes and found SVM.	1. More accurate. 2. Rating can be done with lack of direct information.	1.Application in other field need future investigation
Recommender System for Sports Videos Based on User Audiovisual Consumption	In this authors propose recommender system for game videos, transferred over the web.	1. This keeps away from the concurrence, computation and security issue of central server	1. Particular video fragment can't be recommended using this approach

		approaches in scenarios with a large number of users.	
Travel Recommendation by Mining People Attributes and Travel Group Types From Community-Contributed Photos	In this authors propose a probabilistic personalized travel recommendation model.	1. It gives more satisfactory Results. 2. It improve group recommendations by mining travel preferences	1.The more competitive recommendation models need to be investigated

3. CONCLUSION

Sentiment Analysis is a machine learning issue that has been a research concern for recent years. Although several notable works have come in this field, a completely automated and profoundly well-organized system has not been presented till now. This is because of the unstructured nature of natural language. Latest research shows the use of Sentimental Analysis in developing of the more precise recommender system. These types of approaches are usually used in the e-commerce business. In this paper, we have classified various approaches of the recommender system that are based on a Sentimental analysis.

Future research will deal with advancing the usual techniques and algorithms to enhance the nice of recommender structures predictions and hints. In prospect, more research is required on additional improving the performance measures. Sentiment analysis knows how to be useful for different applications. Though the techniques and algorithms used for sentiment analysis are proceeding fast, on the other hand, many difficulties in this field of study persist unresolved. The highest challenging features exist in use of further languages, dealing by means of negation expressions; produce an instantaneous opinion based on product attributes, the difficulty of sentence/document, handling of inherent product features, etc. Further future research could be devoted to these challenges.

4. ACKNOWLEDGMENTS

We are thankful to the faculty members of School Of Computer Science and Engineering Department of KIIT University, Bhubaneswar for their cooperation and suggestions.

5. REFERENCES

[1] WebLink
https://en.wikipedia.org/wiki/Recommender_system

[2] Sandra Garcia Esparza, Michael P. O'Mahony, Barry Smyth, Mining the real-time web: A novel approach to product recommendation, Elsevier Journal - Knowledge-Based Systems 29 (2012) 3–11.

[3] Verma, Jai Prakash, Bankim Patel, and Atul Patel. "Big data analysis: recommendation system with Hadoop framework." Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on. IEEE, 2015.

[4] Burke, Robin. "Hybrid recommender systems: Survey and experiments." User modeling and user-adapted interaction 12.4 (2002): 331-370.

[5] Shrote, Khushboo R., and A. V. Deorankar. "Review based service recommendation for big data." Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2016 2nd International Conference on. IEEE, 2016.

[6] Krishna, P. Venkata, et al. "Learning automata based sentiment analysis for recommender system on cloud." Computer, Information and Telecommunication Systems (CITS), 2013 International Conference on. IEEE, 2013

[7] Shunmei Meng, Wanchun Dou, Xuyun Zhang, Jinjun Chen," KASR: A Keyword-Aware Service Recommendation Method on Map Reduce for Big Data Applications" IEEE Transactions On Parallel And Distributed Systems, TPDS-2013-12-1141.

[8] G. Kang, J. Liu, M. Tang, X. Liu and B. cao, "AWSR: Active Web Service Recommendation Based on Usage History," 2012 IEEE 19th International Conference on Web Services (ICWS), pp. 186-193, 2012.

[9] Zibin Zheng, Xinmiao Wu, Yilei Zhang, Michael R. Lyu, Fellow, and Jianmin Wang," QoS Ranking Prediction for Cloud Services" IEEE Transactions On Parallel And Distributed Systems, Vol. 24, No. 6, June 2013.

[10] Meghana Ashok, Swathi Rajanna and Pradnyesh Vineet Joshi "A Personalized Recommender System using Machine Learning based Sentiment Analysis over Social Data." Conference on Electrical, Electronics and Computer Science IEEE 2016

[11] X. Yang, Y. Guo, Y. Liu, "Bayesian-inference based recommendation in online social networks," IEEE Transactions on Parallel and Distributed Systems, Vol. 24, No. 4, pp. 642-651, 2013

[12] ANTO, MENARA P., KERALA THRISSUR, MEJO ANTONY, KM MUHSINA, NIVEA JOHNY, VINAY JAMES, and ASWATHY WILSON. "PRODUCT RATING USING SENTIMENT ANALYSIS" International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016

[13] M. Alduan, F. Alvarez, J. Menendez, and O. Baez, "Recommender System for Sports Videos Based on User Audiovisual Consumption,"IEEE Transactions on Multimedia, Vol. 14, No.6, pp. 1546-1557, 2013.

[14] Yan-Ying Chen, An-Jung Cheng, "Travel Recommendation by Mining People Attributes and Travel Group Types From Community-Contributed Photos" IEEE Transactions on Multimedia, Vol. 15, No. 6, October 2013

[15] Kuat Yessenov, Sasa Misailovic, "Sentiment Analysis of Movie Review Comments", 6.863 Spring 2009 final project, pp.1-17

[16] Raghuvanshi, Neha, and J. M. Patil. "A Brief Review of Sentiment Analysis." in International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016

[17] Vohra, Mr Saifee, and Jay Teraiya. "Applications and challenges for sentiment analysis: A survey." International Journal of Engineering Research and Technology. Vol. 2. No. 2 (February-2013). ESRSA Publications, 2013.