

Artificial Intelligence Techniques For Sentiment Analysis of Online Reviews

Aashish Jain, Dr. Arun Malik, Dr. Isha, Vikas Verma

School of Computer Science and Engineering, Lovely Professional University, Punjab

Abstract –Nowadays people are used to express their feelings, their opinions about a topic through comments, tweets or images on social media like twitter and facebook. Social media is widely used by people mostly for this purpose. Sentiment analysis is a process by which we can analyze their comments, tweets, their emotions what they feel about a particular topic by applying machine learning techniques, natural language processing, fuzzy logic based rules etc. In this paper, various techniques of artificial intelligence used for analyzing the sentiments from 2012 to 2019 based on machine learning techniques, natural language processing, deep learning and fuzzy rule based system are discussed.

Keywords - Sentiment Analysis, machine learning, fuzzy logic, random forest, naive bayes, decision tree, support vector machine.

1 INTRODUCTION

Due to increase in the use of internet over past decade, information on the internet is also increasing. There is a large amount of unstructured data available on internet, mostly in text format. For eg- customer's feedback on products and services. Sentiment Analysis provides the ability to capture customer's emotion towards the products and services related to the organization. Reviews are valuable for people who are going to purchase some product. Reviews will help them to better decision. Sentiment analysis face some challenges like extract the information from Amazon, Flipkart, Twitter etc, and then transforming the extracted information into information which . Accurate modeling of topic and Documentation are additional challenges in sentiment analysis. Since sentiment analysis is a supervised machine learning problem, we will be using various supervised machine learning algorithms like Decision tree, Naïve Bayes, Random Forest, Logistic Regression, Support Vector Machine and also some of the most

advanced deep learning techniques like Recurrent Neural Network[3]. By machine learning, we can improve efficiency and accuracy over time which gives the algorithm more experience which is beneficial as in future it can be used to make better predictions and decisions. Moreover, benefit for analyzing product reviews will be beneficial to people. Nowadays, before purchasing any product people are in a habit of checking reviews related to product which are given by the people who are already using that product[15]. Machine Learning and deep learning techniques are discussed below.

Decision Tree is a Supervised Machine Learning technique in which you explain what will be the input and the output corresponding to the data which is trained. In this technique data is splitted continuously according to a specific parameter. In this decision tree, we have two attributes that is parent nodes and leaves where leaves are the result and the parent nodes, where the data is being splitted. It is considered to be the best supervised learning algorithm due to which it is

mostly used. It is basically a predictive model which provides us high accuracy, stability and ease of interpretation. Decision tree is very good at mapping non-linear relationship[4][5]. Terminologies present in decision tree are root node, splitting, decision node, leaf node (Node which do not split), pruning, branch tree, parent and child node.

Pros and cons of decision tree are It is a layman algorithm, Person from non-analytical background can also implement this algorithm, It is the fastest way to significant variables and relation between multiple variables, It can deal with outliers and missing values, It can handle both numerical and categorical values, In case of small data set decision tree leads to over fitting, In case of continuous features decision tree does not performs well.

Logistic regression is a type of classification algorithm. It gives the result in the form of binary outcome i.e. 1 or 0, True or False, Yes or No. It is a special case of linear regression, in linear regression our target outcome is continuous while in logistic regression our target outcome is categorical. It is used for prediction of outcome of dependent variable which is based on more than one independent variables (predictor). It is also used to detect if each feature is meaningful or not. It is a technique which is statistical which is used for calculating the relations between the variables[1][3]. This model recognizes features and calculate the coefficient and weight for independent feature and then computes the category of tweet as a vector[2].

2 RELATED WORK

Pictures on social media are provided with a caption under it which is related to the image. In 2019, Feiran Huang and Xiaoming Zhang et al. [31] proposed a deep multimodal attention Fusion(DMAF) which will analyze both text and

visual emotions present in the image and it will also find the correlation between the text and the visual emotion. They used single visual model, single textual model, early fusion, late fusion, VAM, SAM, MAM, DMAF techniques to do sentiment analysis of image and text and got 86.9% accuracy using DMAF method.

People nowadays are in a habit of sharing their emotions, feelings, opinions about a specific topic on a social media platform like twitter, facebook etc. Mostly they share their thoughts in the form of text. In 2019, Kashfia Sailunaz Reda Alhajj[29] extracted twitter tweets and comments on a specific topic and proposed a model which will detect public emotions which they expressed in terms of text i.e. tweets and they have analyzed sentiments and used them for recommending who all are agreed to this topic and have expressed positive feelings for that. They have used Naïve Bayes, SVM, Random Forest for analysis and got 66.86% accuracy using naïve bayes.

Over the past few years, deep learning is mostly used for image classification, speech recognition etc but it is somewhere limited and it is observed in few cases that existing models for analysis of sentiments are unable to extract some implicit aspects. In 2019, Paramita Ray and Amlan Chakrabarti [25] proposed a deep learning model to extract the aspects from the text and analyze the user's sentiment related to the aspect. They used CNN and core NLP for analysis.

In 2019, Erick Kauffmanna and Jesús Peralb et al. [23] proposed a framework which can detect fake reviews and automatically analyze the user generated reviews about the product, brands in positive or negative opinions and after that it will automatically rate the product or brand based on that analysis. They used Natural Language processing and got 85.3% accuracy in finding fake reviews.

People share their thoughts, their opinions like tweets on twitter, comments on facebook which is best source for sentiment analysis. In 2018 Megha Rathi and Aditya Malik et al. [22] Proposed a hybrid machine learning approach by merging decision tree with support vector machine to increase efficiency and reliability. They used SVM, Adaboost D-Tree, Decision Tree got 84% accuracy using decision tree.

In 2017, Anees Ul Hassan and Jamil Hussain et al. [21] proposed a model by using natural language processing techniques and machine learning techniques to calculate the depression level of a user by extracting and observing emotions from the text present on different social media platforms. They used Naïve Bayes, Support Vector Machine and got 91% accuracy using support vector machine.

To grow the restaurant business most important thing is the customer feedback. Through which they get to know what is the best part and what are the things they need to work on in their restaurant. In 2016 Tri Doan and Jugal Kalita [19] by using random forest i.e. machine learning technique created a model to perform sentiment analysis on customer reviews.

Due to increase in quantity of the content generated by the users on internet not only in English but in different languages. Natural language processing can be used for task dealing and classification from texts which is also a part of sentiment analysis. Supervised machine learning algorithms can be used to build models to detect and classify sentiments in different languages. In 2013, Alexandra Balahura and Marco Turchib [10] proposed a model using bagging technique for German for which accuracy score is 65.9%.

Product reviews are very helpful for customer to make better decisions about the product to purchase or not. In 2017 Wei Zhao and Ziyu

Guan et al. [9] designed a deep learning framework for product review classification using Lexicon, SVM, NBSVM, CNN-weak and many more techniques. Using NBSVM authors got 82.6% accurate.

As the data is produced very fast due to increase in the use of social media, people uses this data for their future decisions due to which automatic summarization of opinion of the users is needed. It can be done using artificial intelligence algorithms like Naïve Bayes, maximum entropy, SVM etc. summarization is done on the customer review of the product. In 2017 Rajalaxmi Hegde and Dr. Seema[8] proposed a model using SVM and Naïve Bayes to increase the accuracy score. As a result, Proposed method increases accuracy to 83% . Naïve Bayes gives accuracy 78.44% and SVM gave 80.34%.

SaaS (Software as a Service) is frequently used for software delivery models. Due to increase in Edge computing, it becomes difficult to select the best SaaS service provider. Amazon Web Services and Microsoft provide different frameworks and service quality pillars for cloud applications. To overcome this drawback, a sentiment approach has been done in 2019 by Muhammad Raza and Farookh Khadeer Hussain[7] to predict which service quality pillar they belong to. A comparative analysis is done to identify which machine learning classification model best performs on SaaS reviews. They used KNN, Passive aggressive, Logistic Regression, Decision Tree, Ridge, Linear SVM etc techniques.highest Accuracy score achieved was 85% using Logistic Regression.

People are nowadays in a habit of expressing their views in terms of picture and text. In 2019, Ziyuan Zhaoa and Huiying Zhua et al. [26] proposed a technique which will find the relation between the text and the image. They used SentiBank, Visual model, textual model for analyzing the sentiments and got 84% accuracy

using sentibank approach. Table 1 describes the literature review for the research work done in the field of artificial intelligence so far.

Table 1: Literature Review

S.No.	Author Name	Year	Technique used	Result
1	Farookh Khadeer Hussain, Muhammad Raza, Omar Khadeer Hussain, Ming Zhao, Zia ur Rehman [7]	2019	KNN, Passive aggressive, Logistic Regression, Decision Tree, Ridge, Linear SVM	Accuracy score obtained is 85% using Logistic Regression
2	Kariman Elshakankery , Mona F. Ahmed [12]	2019	RNN, SVM	highest accuracy score achieved is 56.17% using RNN technique.
3	Rania Hodhod, Mohammed H. Abd El-Jawad, Yasser M. K. Omar [20]	2019	CNN, Random Forest, Decision Tree, Naïve Bayes	accuracy score is 83.7%
4	Erick Kauffmanna, Jesús Peralb, David Gilc, Antonio Ferrándezb, Ricardo Sellersd, Higinio Morac [23]	2019	Natural Language Processing	85.53% accurate in finding fake reviews
5	Mandar Gogate, Jingpeng Li, Kia Dashtipour, Amir Hussain, Fengling Jiang, Bin Kong [24]	2019	Suort Vector Machine, Logistic Regression, DNN and CNN classifier	85.91% accurate using CNN
6	Paramita Ray, Amlan Chakrabarti [25]	2019	CNN, Core NLP	accuracy score is 87%
7	Ziyuan Zhaoa, Huiying Zhua, Zehao Xuea, Zhao Liua, Jing Tiana, Matthew Chin Heng Chuua, Maofu Liu [26]	2019	SentiBank, Visual model, textual model	84% accurate using sentibank approach
8	Luoyi Fu , Bin Yao , Feilong Tang , Wenchao Xu [27]	2019	MaxEnt JABST	Imroves accuracy by 5%
9	Akhilesh Dwivedi, RP Pant [28]	2019	ETRSSS, WTRSSS (Weighted ternary reduced soft sentiment set)	ETRSSS gives better results than WTRSSS
10	Kashfia Sailunaz Reda Alhajj [29]	2019	Naïve Bayes, SVM, Random Forest	66.86% accuracy using naïve bayes

11	Srishti Vashishtha, Seba Susan [30]	2019	Fuzzy logic, SVM	fuzzy rule based performs best and gives best results as compared to SVM
12	Xiaoming Zhang, Feiran Huang, Jie Xu, Zhonghua Zhao, Zhoujun Li [31]	2019	Early fusion, late fusion, VAM, SAM, MAM, DM AF	86.9% accuracy using DMAF method
13	Tomas Kincl, Michal Novak, Jiri Pribil [32]	2019	logistic Regression, Naïve Bayes	79.15% accurate
14	Fang Chen, Yongfeng Huang [33]	2019	LSTM, AT-LSTM, MemNet	accuracy score is 92%
15	Yang Liu, Jian-Wu Bi, Zhi-Ping Fan [34]	2019	Fuzzy logic	Accuracy score improved
16	Xiaoming Zhang, Senzhang Wang, Chaozhao Li, Zhoujun Li, Yueying He [35]	2019	Early Fusion, late fusion, CCR, HDF	85.9% accurate using HDF
17	Xiaofei Xu, Xutao Li, Xiaojun Chen, Bowen Zhang [6]	2019	Text-CNN, CNN+rule	proposed model gives 97.71% accuracy
18	Asad Abdi, Siti Mariyam Shamsuddin, Shafaatunnur Hasan [15]	2018	SVM, Logistic Regression	26.35% using SVM
19	Charu Nanda, Mohit Dua and Garima Nanda [17]	2018	Naïve Bayes, SVM, Random Forest	89.73% accuracy score using SVM
20	Rachana Bandana [18]	2018	Linear SVM, Naïve Bayes	84% using Naïve Bayes
21	Aditya Malik, Daksh Varshney, Megha Rath [22]	2018	SVM, Adaboost D-Tree, Decision Tree	84% using decision Tree
22	Rajalaxmi Hegde, Dr. Seema [8]	2017	Naïve Bayes, SVM	Proposed method increases accuracy to 83%
23	Ziyu Guan, Wei Zhao, Long Chen, Deng Cai, Beidou Wang, Xiaofei He, Quan Wang [9]	2017	Lexicon, SVM, NBSVM, CNN-weak	Using NBSVM accuracy score is 82.6%
24	Omar Qawasmehb, Mohammad AL-Smadia, Yaser Jararweha, Mahmoud Al-Ayyoub, Brij Gupta [11]	2017	RNN, SVM	Its accuracy is 95.4%

25	Marouane Birjalia, Abderrahim Beni-Hssanea, Mohammed Erritalib [13]	2017	Naïve Bayes, SMO	82% accuracy using naïve bayes
26	Arman S. Zharmagambetov, Alexandr A. Pak [16]	2017	Recurrent Neural Network	89.8% accurate
27	Jamil Hussain, Anees Ul Hassan, Muhammad Sadiq, Sungyoung Lee, Musarrat Hussain, [21]	2017	Support Vector Machine, Naïve Bayes	obtained accuracy score is 91% using Support Vector Machine
28	Tri Doan, Jugal Kalita [19]	2016	Random Forest	accuracy score is 92%
29	Alexandra Balahura, Marco Turchib [10]	2013	SMO, Adaboost MI, Bagging	using bagging technique accuracy score is 65.9%
30	Liam Donaldson, Daniel Ramirez-Cano, Christopher Millett, Ara Darzi, Felix Greaves[14]	2012	logistic Regression, Decision Tree, Spearman test rank correlation	89% overall recommendation

3 CONCLUSION

In this paper, various artificial intelligence techniques like deep learning, machine learning, and natural language processing and fuzzy logic are discussed. The last seven years studies on sentimental analysis for online reviews are described in this paper. This paper will help novice to select the best techniques used for sentiment analysis of online reviews.

REFERENCES

- [1] Jonghee Kim, Jonghwan Lee, Chungsu Lee, Eunsoo Park, Junmin Kim, Hakil Kim, Jaeun Lee, Hoeri Jeong "Optimal Feature Selection for Pedestrian Detection based on Logistic Regression Analysis" : 2013 IEEE International Conference on Systems, Man, and Cybernetics
- [2] B. Pang, L. Lee, S. Vaithyanathan, "Thumbs up?: Sentiment Classification Using Machine Learning Techniques." In conference on Empirical methods in natural language processing in association for computational Linguistics, Vol.10, pp. 79–86, 2002.
- [3] Zohreh Madhoushi, Abdul Razak Hamdan, Suhaila Zainudin "Sentiment Analysis Techniques in Recent Works" Science and Information Conference 2015
- [4] Zhang Xiang, "The Research of Decision Tree Classification Algorithm in Data Mining" Zhenjiang: Jiangsu Science and Technology University, 2008.
- [5] Anuradha, Dr. Gaurav Gupta, "A Self Explanatory Review Of Decision Tree Classifiers", IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), 2014
- [6] Bowen Zhang , Xiaofei Xu , Xutao Li, Xiaojun Chen, Yunming Ye, Zhongjie Wang, " Sentiment analysis through critic learning for optimizing convolutional neural networks with rules", Neurocomputing 356 (2019)
- [7] Muhammad Raza, Farookh Khadeer Hussain, Omar Khadeer Hussain, Ming Zhao, Zia ur Rehman, "A comparative analysis of machine learning models for quality pillar assessment of SaaS services by multi-class text classification of users' reviews" , Future Generation Computer Systems 101 (2019)
- [8] Rajalaxmi Hegde, Dr. Seema, "Aspect Based Feature Extraction and Sentiment Classification of Review Data sets using Incremental Machine learning Algorithm", 3rd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB17), 2017
- [9] Wei Zhao, Ziyu Guan , Long Chen, Xiaofei He, Deng Cai, Beidou Wang, Quan Wang, "Weakly-supervised Deep Embedding for Product Review Sentiment Analysis", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, AUGUST 2017

- [10] Alexandra Balahura, Marco Turchib, "Comparative experiments using supervised learning and machinetranslation for multilingual sentiment analysis", Computer Speech and Language xxx (2013)
- [11] Mohammad AL-Smadia, Omar Qawasmehb, Mahmoud Al-Ayyouba, Yaser Jararweha, Brij Gupta,"Deep Recurrent Neural Network vs. Support Vector Machine for Aspect-Based Sentiment Analysis of Arabic Hotels' Reviews", 2017
- [12] Kariman Elshakankery , Mona F. Ahmed, "HILATSA: A hybrid Incremental learning approach for Arabic tweets sentiment analysis", Egyptian Informatics Journal, 2019
- [13] Marouane Birjalia, Abderrahim Beni-Hssanea, Mohammed Erritalib, "Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks", The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017)
- [14] Felix Greaves, Daniel Ramirez-Cano, Christopher Millett, Ara Darzi, Liam Donaldson, "Machine learning and sentiment analysis of unstructured free-text information about patient experience online", 2012
- [15] Asad Abdi , Siti Mariyam Shamsuddin , Shafaattunnur Hasan, "Machine learning-based Multi-Documents Sentiment-oriented Summarization using Linguistic treatment", 2018
- [16] Arman S. Zharmagambetov, Alexandr A. Pak, "Sentiment Analysis of a document using deep learning approach and decision trees", 2017
- [17] Charu Nanda, Mohit Dua and Garima Nanda, "Sentiment Analysis of Movie Reviews in Hindi Language using Machine Learning", International Conference on Communication and Signal Processing, April 3-5, 2018, India
- [18] Rachana Bandana, "Sentiment Analysis of Movie Reviews Using Heterogeneous Features", IEEE,2018
- [19] Tri Doan, Jugal Kalita, "Sentiment Analysis of Restaurant Reviews on Yelp with Incremental Learning", 2016 15th IEEE International Conference on Machine Learning and Applications
- [20] Mohammed H. Abd El-Jawad, Rania Hodhod, Yasser M. K. Omar, "Sentiment Analysis of Social Media Networks Using Machine Learning", 2019
- [21] Anees Ul Hassan, Jamil Hussain, Musarrat Hussain, Muhammad Sadiq, Sungyoung Lee, "Sentiment Analysis of Social Networking Sites (SNS) Data using Machine Learning Approach for the Measurement of Depression", IEEE, 2017
- [22] Megha Rathi, Aditya Malik, Daksh Varshney, Rachita Sharma, Sarthak Mendiratta, "Sentiment Analysis of Tweets using Machine Learning Approach", Proceedings of 2018 Eleventh International Conference on Contemporary Computing (IC3), 2-4 August, 2018, Noida, India
- [23] Erick Kauffmanna, Jesús Peralb, David Gilc, Antonio Ferrándezb, Ricardo Sellersd, Higinio Morac, "A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making", Industrial Marketing Management, 2019
- [24] Kia Dashtipour, Mandar Gogate, Jingpeng Li, Fengling Jiang, Bin Kong, Amir Hussain, "A Hybrid Persian Sentiment Analysis Framework: Integrating Dependency Grammar Based Rules and Deep Neural Networks", Neurocomputing, 2019
- [25] Paramita Ray, Amlan Chakrabarti, "A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis", Applied Computing and Informatics, 2019
- [26] Ziyuan Zhaoa, Huiying Zhua, Zehao Xuea, Zhao Liua, Jing Tiana, Matthew Chin Heng Chuua, Maofu Liu, "An image-text consistency driven multimodal sentiment analysis approach for social media", Information Processing and Management 56 (2019) 102097
- [27] Feilong Tang , Luoyi Fu , Bin Yao , Wenchao Xu, "Aspect based fine-grained sentiment analysis for online reviews", Information Sciences, 2019
- [28] Akhilesh Dwivedi, RP Pant, "An Algorithmic Implementation of Entropic Ternary Reduct Soft Sentiment Set (ETRSSS) using Soft Computing Technique on Big Data Sentiment Analysis (BDSA) for Optimal Selection of a Decision based on Real-Time Update in Online Reviews", Journal of King Saud University - Computer and Information Sciences, 2019
- [29] Kashfia Sailunaz Reda Alhaji, "Emotion and Sentiment Analysis from Twitter Text", 2019
- [30] Srishti Vashishtha,Seba Susan, "Fuzzy Rule based Unsupervised Sentiment Analysis from Social Media Posts", Expert Systems With Applications, 2019
- [31] Feiran Huang, Xiaoming Zhang, Zhonghua Zhao, Jie Xu, Zhoujun Li, "Image-text sentiment analysis via deep multimodal attentive fusion", Knowledge-Based Systems, 2019
- [32] Tomas Kincl,Michal Novak,Jiri Pribil, "Improving sentiment analysis performance on orphologically rich languages: Language and domain independent approach", Computer Speech&Language56(2019)36_51
- [33] Fang Chen , Yongfeng Huang, "Knowledge-enhanced neural networks for sentiment analysis of Chinese reviews", Neurocomputing, 2019
- [34] Jian-Wu Bi, Yang Liu, Zhi-Ping Fan, "Representing sentiment analysis results of online reviews using interval type-2 fuzzy numbers and its application to product ranking", Information Sciences 504 (2019) 293–307
- [35] Jie Xu, Feiran Huang, Xiaoming Zhang, Senzhang Wang, Chaozhua Li, Zhoujun Li, Yueying He, "Sentiment analysis of social images via hierarchical deep fusion of content and links", Applied Soft Computing Journal, 2019