SENTIMENT ANALYSIS USING MACHINE LEARNING: A REVIEW

Anjali Gupta[1], Amita Dhankar[2], Surayansh Dabas[3]
M.Tech student, Assistant Professor, B.Tech MSIT Delhi
Department of C.S.E.
UIET, M.D. UNIVERSITY, Rohtak, Haryana India

Abstract: Sentiment analysis is a field of Research under analytics. This can use computational techniques by reading raw data to make sense. This is called analysis. By using sentiment analysis written expression can be evaluated which can be favorable, unfavorable or neutral. People use different types of social media like facebook, twitter, people sentiments can be determined in an effective way by twitter. In this various machine learning techniques are used. Raw data is read from twitter and it is compared against trained machine which can be positive, negative or neutral. Social networking data can be examined by effectual sentiment analysis which involves extrication of subjective information from written data.

Keywords: Twitter, KNN, SVM, Crisp-DM, and SAS-SEMA.

Introduction

Over last years, sentiment category has made vital interest. Large amount of facts generated on line can be processed into data to help managerial, operational and decision making. Sentiment analysis also seems to be developing in non topical and text evaluation. Opinion mining is also same as sentiment analysis in which out of the supplied sentences we can determine which one is positive, negative, and impartial. Another function it performs is that it identifies and extracts reviews and attitudes from the given text.

Machine Learning

It is the field of computer science that provides ability to computers to learn without programming explicitly. It is also used insentiment analysis in which we can identify that the text generated by the user is positive, negative or neutral. Machine learning generally differentiates three types of learning- supervised, weakly supervised and unsupervised learning. Reinforcement learning also comes under machine learning technique but is it is not used to classify the text.

Supervised Learning

Supervised learning consists of an input value and the desired output value.

NAIVE BAYES Most popular algorithm used for the classification of the text. Naïve Bayes classifier finds out the probability of different attributes of data which a certain class consists of. It is based on Bays theorem.
P(A/B)= (B/A),P(A)/P(B)
It assumes that the combining probability is independent of each other.

Maximum Entropy is characterized by the set of weights. It is also known as the exponential classifiers because it works by getting some features from the input and then combine them linerally and then used as the sum of exponent.

Decision tree It is a type of classifiers in tree in which the internal nodes represent features, edges represent tests and the leaves represent categories. It starts processing from the root of the tree and ends until the goal is reached. It is used in speech and language processing.

KNN Algorithm or Memory based learning In this algorithm the object is classified by maximum vote of its neighbor and the object classified is most common amongst its K neighbors. It is used to determine the similarity between the already and the new data by looking in the memory.

Support Vector Machines These are used to get results with higher accuracy and it operates by constructing a hyper plane which consists of maximum Euclidean distance to the training examples.

Unsupervised Learning It involves the learning patterns in the input in which there are no output values supplied. Some of the common examples of this are K mean clustering or cluster analysis.

LITERATURE REVIEW

B.LangLi2002
This paper determines the user sentiments on different emotions and emotions were classified on the basis of their frequency of use and popularity. Depending upon the frequency and popularity, weights were assigned to emotions.[9]

Daiyanet.al. 2003
This paper linked the changes in the stock price of Hathaway group when affected by anna Hathaway positively which are although not related. Further sentiment classification wasoffered to assign positive or negative class to the documents but fails to find out whether likedor disliked by the reviewers.[4]

A.Shoukry's 2005
This paper is on explaining utility of Arabic sentiments analysis for Arabic tweets. In thispaper purpose was to identify whether a sentence is weblog,review,tweet or any other.[3]

Rappaport's 2008
This paper clearly shows importance of sentiment analysis. It uses digital metrics field guideand various workshops were conducted.[28]

Saitta 2010
In this importance of MAT lab has been studied for sentiment analysis. It gives us
knowledge how to use MAT lab more effectively and statically so that it would be added tomachine learning. In thispaper blog by loren was used on MAT lab and twitter was used for
analyzing.[26]

Krzysztof et.al.2011
In this paper KNN algorithm has been studied to measure performance and find out result by
plotting andfinding that whether the collected data is neutral or swaying on one side.[29]

BingLiu'set. al. 2012
In this paper natural language mining concepts have been studied. In this different features of opinion miningare explained. The basic idea was to work in two main areas, first one is mining opinions for e.g. this product is best and other is mining comparative opinions such as this product is better than that product.[15]

HemLatha et. al. 2013
In this machine learning algorithm was used to perform supervised learning which considersemotions as noisy labels. In this Naive Bayes and Maximum entropy classification wereused to achieve high accuracy.[10]

Khairnaret.al. 2013.
Naïve bays and SVM techniques are discussed. In this paper main focus is on sentiment classification with the help of existing approaches by using machine learning techniques. It
concludes that all the sentiment tasks are very challenging and SVM is the best method fordata classification.[14]

S.Padmajaet. al. 2014 .
They did research to find out scope of negation in newspaper. In this similar techniques are used to find out the sentiment conclusion. For e.g. "Abki bar ModiSarkar" it is a positive phrase and "Abki bar NahiChahiyeaiseSarkar" this is a negative phrase. Both of them possess different sentiments and identified with words like Nahi.[27]

Riley's et.al. 2015 .
It works to see how it works with large sets of data. He did research for Meta data plannng and for digital library team i.e. production -oriented. He combines different sentiments and did research to find out how the conclusion of one sentiment effects the other.[5]

Piatetsky et. al. 2015 .
In this various models have been studied. Models like Crisp-DM and SAS SEMA has been studied for the analysis of Sentiment analysis.[7]

Kharde et. al. 2016.
In this paper a survey is provided and different existing techniques are compared for opinion mining such as machine learning, Lexicon based approaches. In this research it shows that SVM and Naïve Bayes provide high accuracy. We conclude that higher accuracy can be achieved by cleaner data.[8]

Pruthi et.al. 2017 .
In this paper SVM and KNN model were used to improve the classification accuracy. In this tweets were analyzed whether they are positive, negative or neutral sentiments.[21]

Gupta et.al. 2017 .
This paper focuses on analyzing sentiments by putting the data in machine learning model then train it and accuracy is checked. It includes data collection, text preprocessing and sentiment detection, classification, training and at last testing. But its disadvantage is that it doesn't comprise of dimension of diversity in data.[24]

Kawade et.al. 2017 .
In this an example of terrorist attack was taken and after that people views are preprocessed by different techniques to determine the polarity as well as the emotions. It defined that in future big data analysis technique will be used to classify emotions.[14]
PalakBaid et.al. 2017
In this different techniques like Naive Bayes, K-nearest neighbours, random forest were used to identify polarity of tweets. Best results were achieved By Naive Bayes. It helps in various domains.[20]

Conculsion

Investigation was done on informational indexes of expressions and suppositions were appointed to them. Weights were doled out to words with prevalence .These words were contrasted against phrases with alter the net estimation proportion and conclusion was accomplished if a general tweet from twitter has a positive feeling or negative slant. In this various techniques like Naive Bayes, KNN, Maximum Entropy and SVM are used.

References

[1] Bing Liu, 2010, "Sentiment analysis and subjectivity", Handbook of natural language mining.

[2] R. Mukras, J. Carroll ,2004, "A comparison of machine learning techniques applied to sentiment classification" , Indian Journal of Computer Science and Engineering (IJCSE) pp 2000-2004

[3] A. Shoukry, A. Rafea, 2012, "Sentence- Level Arabic Sentiment Analysis", 978-1-4673-1382, IEEE . Available: http://ieeexplore.ieee.org/abstract/document/6261103/?reload=true.

[4] Md. Daiyan, Dr. S.K.Tiwari , 4, April 2015, "A literature review on opinion mining and sentiment analysis", International Journal of Emerging Technology and Advanced Engineering, Volume 5.

[5] Jenn Riley, January 18 2017, "Understanding Meta Data", Primer Publication of National Information Standard Organization .Baltimore . [11].

[6] S Padmaja and Prof S Sameen Fatima, 17 December 2016, "Evaluating Sentiment Analysis: Identifying Scope of Negation in Newspaper Articles", ,UCEOsmania University . IJARAI .

[7] Gregory Piatetsky, 10 October 2014, "Why CRISP_DM model is the Many popular methodology for data analytics". Available: http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html.

[8] "Use of SVM for Binary Classification", January 2017, Mat lab stats., Math works. Available: https://in.mathworks.Com/help/stats/support-vector-machines-for-binary-classifcation.html?requestedDomain=www.mathworks.com. By Vishal

[9] Bo Pang and Lillian Lee, 2008, "Opinion mining and sentiment analysis", Foundations and trends in information retrieval, Vol. 2, No 1-2 (2008) 1–135.

[10] I.Hemalatha1, March – April 2013 ,"Sentiment Analysis Tool using Machine Learning Algorithms", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 2, Issue 2, , ISSN 2278-6856

[11] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification," in Proc. 14th ACM Int. Conf. Inf. Knowl.Manage., 2005, pp. 617–624.

[12] V. N. Vapnik, The Nature of Statistical Learning Theory.NewYork:Springer-Verlag, 1995.

[13] "Clustering High Dimensional Data Using SVM"Tam P. Ngo,December 2006.

[14] M.Govindarajan, Sentiment Analysis of Movie Reviews using Hybrid Method of Naive Bayes and Genetic Algorithm , International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970), Volume-3 Number-4 Issue-13 December-2013 by Dipak and Jayashri.

[15] Mukherjee, Arjun and Bing Liu. Aspect Extraction through SemiSupervised Modeling.inroceedings of 50th Anunal Meeting of Association for Computational Linguistics (ACL-2012) (Accepted for publication). 2012.

[16] Mukherjee, Arjun and Bing Liu. Modeling Review Comments.in Proceedings of 50th Anunal Meeting of Association for Computational Linguistics (ACL-2012) (Accepted for publication). 2012.

[17] PeimanBarnaghi, John G. Breslin and ParsaGhaffari, "Opinion Mining and Sentiment Polarity on Twitter and Correlation between Events and Sentiment", 2016 IEEE Second International Conference on Big Data Computing Service and Applications.

[18] MondherBouazizi and TomoakiOhtsuki, "Sentiment Analysis: from Binary to Multi-Class Classification", IEEE ICC 2016 SAC Social Networking, ISBN 978-1-4799-6664-19. [5] NehalMamgain, Ekta Mehta, AnkushMittal and Gaurav Bhatt, "Sentiment Analysis of Top Colleges in India Using Twitter Data", (IEEE) ISBN - 978-1-5090-0082-1, 2016.

[19] Hassan Saif, Miriam Fernandez,PalakBaid, Yulan He, and HarithAlani. Evaluation datasets for twitter sentiment analysis: a survey and a new dataset, the sts-gold. 2013.

[20] JurgenSchmidhuber, JyotikaPruthi, Deep learning in neural ¨ networks: An overview. Neural networks, 61:85–117, 2015. .

[21] SoroushVosoughi, Helen Zhou, and Deb Roy. Enhanced twitter sentiment classification using contextual information. arXiv preprint arXiv:1605.05195, 2016.

[22] A. Balahur, J. Hermida and A. Montoyo, 'Building and Exploiting Emotinet, a knowledge base for emotion detection based on the appraisal theory model', Affective Computing, IEEE Transactions, vol. 3, 188101, 2012.

[23] G. Vinodhini, Bhoomika Gupta and R. Chandrasekaran, 'Sentiment analysis and opinion mining: A survey', International Journal, vol. 2, 6, 2012. www.wikipedia.com

[24] SandroSaitta , Jul 13 2007 , "Why is mat lab the best language for data mining" ,Data Mining Research. Available: http://www.dataminingblog.com/why-is-matlab-the-best-language-for-data-mining/

[25] S Padmaja and Prof S Sameen Fatima, 17 December 2016, "Evaluating Sentiment Analysis: Identifying Scope of Negation in Newspaper Articles", ,UCEOsmania University . IJARAI .

[26] Nancy Lazarus. 21 July 2015, "5 Key challenges of sentiment analysis", Blog.Ad weeks. Available:

[27] http://www.adweek.com/digital/5-key-challenges-in-sentiment-analysis/.

[28] Krzysztof JĘDRZEJEWSKI2, Maurycy ZAMOR SKI3, 2013, "Performance of K-Nearest Neighbors Algorithm", Foundations of Computing and Decision Science,Vol-38,No.2