

Creation of Hybrid Classifier Algorithm To Improve The Results of Existing Classifier Algorithms For Emotion Analysis on Social Media Content

Savleen Kaur

Assistant Professor, School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India. Pin Code: 144411
iridiumsavleen@gmail.com

Maneet Kaur

Assistant Professor, School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India. Pin Code: 144411
maneet.rainu@yahoo.com

Abstract- Information mining is utilized to find the concealed realities and examples from the mass information for forecast. Using one of the popular classifier methodologies we are performing the emotion analysis on opinion-rich resources like: online review sites and personal blogs etc. to understand the opinions of other people. In this paper we applied the improved classifiers on social media texts collected from Twitter and Facebook.

Keywords - Decision-Tree Based Classification, Bayesian Classifier, Laplacian Correction, Social Media Data Mining, Rule Based Classification

I. INTRODUCTION

Data mining [1, 2] is a technique to automatically discover and extract the hidden or undiscovered information from historical data. Web mining under data mining is further divided into three categories, one is web structure mining second is web content mining and third one is web usage mining. Fig. 1 illustrates the three categories of web mining. Web usage mining understands the user's behaviors by analyzing the historical data.

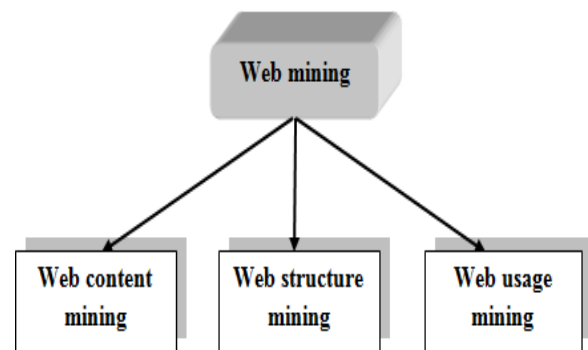


Fig.1. Types of web mining

Now the days we are keen users of social media like: Twitter, Facebook and WhatsApp. Applying data mining classifiers for emotion analysis will help us to understand the opinions of other people. Stream mining is highly significant for opinion mining [3, 4] and several application areas related to prediction in real world. In web data mining [5], stream data is exigent task because it is real data which changes frequently timely. In stream mining immense data is produced online from different sources.

We are sharing our emotions on social media to attract the attention or bully someone. Sentimental analysis [6] is used to address many critical issues nation is dealing now the days. Beyond all other anti-social behavior is very sensitive concern.

II. REVIEW OF LITERATURE

A. Classification

Grouping is a powerful strategy in Data Mining to anticipate clear cut separate and disordered class names [7, 8]. It is a principle issue in Data Mining. An order model can be built to anticipate regardless a specific gathering will win or not. In this paper assignment of information investigation is grouping in which the classifier is developed to foresee downright class name for ideological groups.

General model of Classification process

The working of data classification is interpreted in two steps, as in Fig. 2, comprise of a learning step which is containing former data (construction of classification model), and Fig. 3, a classification model for testing (model to predict class labels for testing data).

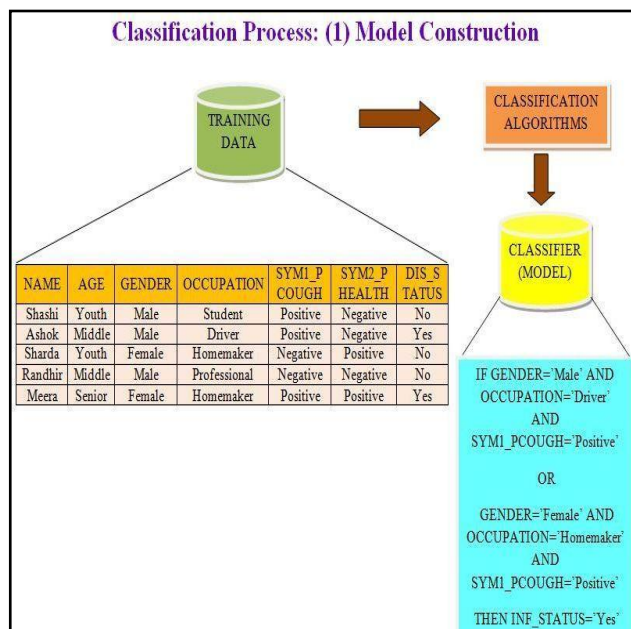


Fig. 2. Classification Learning Model

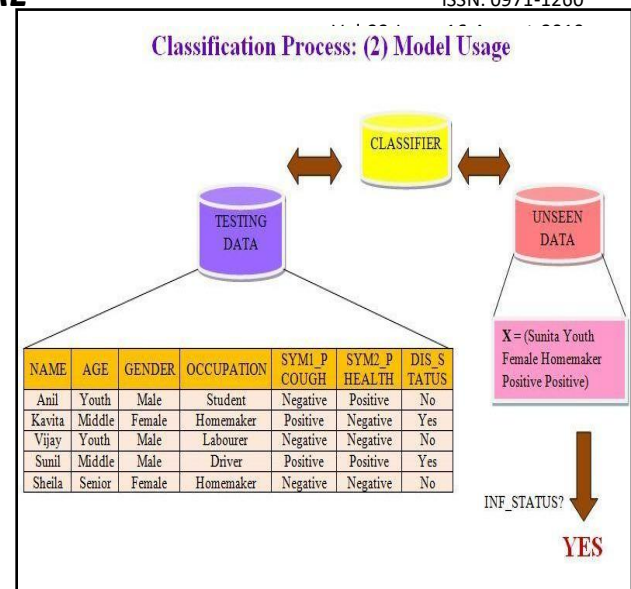


Fig. 3. Classification Testing Model

B. Decision-Tree Classification

As Decision-Tree Classification [9] belongs to the supervised learning and decision tree always learn from trained data. A decision tree is constituted as a tree structure in which top node is always considered as root node and each line/branch represents final outcome of the testing data and each non-leaf node indicates a testing attribute along with condition.

C. Bayesian Classification

Bayesian Classification is based on Bayes' Theorem which is used to predict class membership probabilities [10, 11]. In this paper, we have used Simple Naïve Bayesian Classifier [12] which presumes the outcome of characteristic incentive on a given class which is not dependent on supplementary attributes (*independent class condition*).

III. PROPOSED APPROACH

Using sentiment analysis we are analyzing the social media data in an optimized way. Our hybrid algorithm is generated from the concept of supervised learning classifiers which is used to find out the probability of occurrence of some event on the basis of trained data.

In Fig. 4 representing the working of decision tree

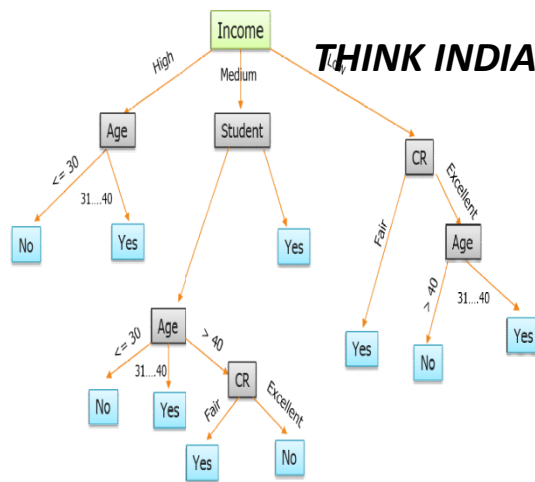


Fig. 4. Decision Tree Example

Steps of research work followed mentioned below:-

- Collection of data.
- Pre-Process (transformation and handling noisy data) the fetched data and remove the punctuation, stop word and repetition from the data.
- Split the data set into different labels i.e. provides the positive or negative label as per the tweets we get from the twitter.
- Applied predictive model of Naïve Bayes to calculate the region wise opinion.
- Applied predictive model of decision tree to classify the given review by the users of different parties.
- System will generate the results on the basis of above mentioned steps.
- Accuracy measurement of work.

IV. MODEL EVALUATION

As India is an equitable nation where we picked our delegate through political decision and which gathering will win the political race based on remarks given by the residents of the nation on twitter. The achievement of the framework relies upon exactness and accuracy esteem. Our exploration model creates the high precision with respect to party win based on individuals feeling.

A predictive model can be evaluated on the basis of mentioned parameters: time efficiency, result

accuracy, memory efficiency etc.

Using text mining algorithm we fetched the frequency of words being used on tweeter, we fetched the data of top ten most frequent words.

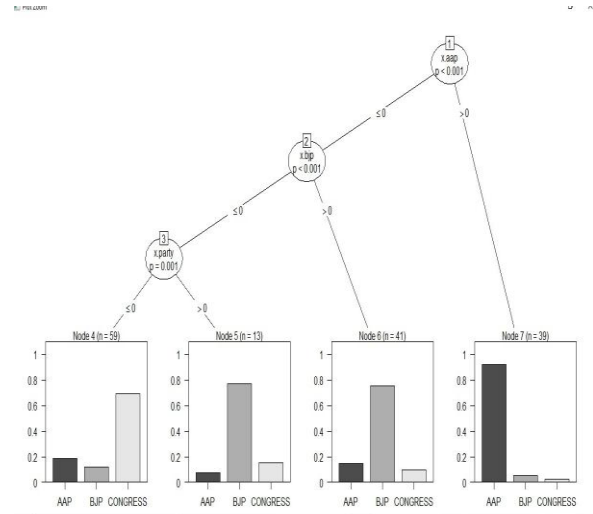


Fig. 5. Decision Tree Model on Election Data

In Fig.5 Applied decision tree predictive model on clean and scaled data, results are represented in the form of tree where four nodes have been generated and concluding which party is having high chances to win the elections. As model describes that the chances of BJP to grab the victory in election is more in other state. But in one place the congress is beating the BJP party.

The applied predictive model defines the accuracy of our hybrid system which is around about 81%.

V. RESULT AND DISCUSSION

Comparison in Fig.6 clearly represents that our system is more accurate and efficient than others.

The comparison of the accuracy, precision1, precision2 and precision3 of the two different classifiers is explained.

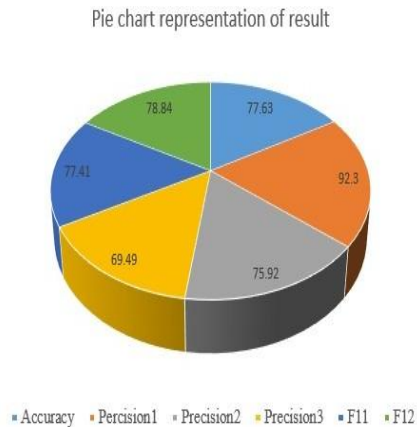


Fig. 6. Accuracy Model

Accuracy generated by hybrid system is superior then naive Bayes. Fig.7 represents the result comparison between naive Bayes and hybrid system.

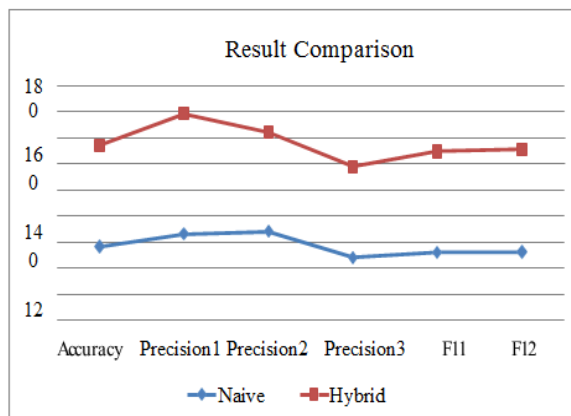


Fig. 7. Comparison of Results

VI. CONCLUSION

Hybrid algorithm is producing highly efficient results considering Reduced completion time, Minimized the error rate and Increase in performance of Analysis, as well as there are lots of work can be done to enhance the algorithm and increase the accuracy as well as the reliability of the system.

REFERENCES

[1] Berkhin P. A survey of clustering data mining techniques. In Grouping multidimensional data 2006 (pp. 25-71). Springer, Berlin, Heidelberg.

[2] Rygielski C, Wang JC, Yen DC. Data mining techniques for customer relationship management. Technology in society. 2002 Nov 1;24(4):483-502.

[3] Liu B, Zhang L. A survey of opinion mining and sentiment analysis. In Mining text data 2012 (pp. 415-463). Springer, Boston, MA.

[4] Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems. 2015 Nov 1;89:14-46.

[5] Johnson F, Gupta SK. Web content mining techniques: a survey. International Journal of Computer Applications. 2012 Jan 1;47(11).

[6] Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval. 2008 Jul 7;2(1-2):1-35.

[7] G. Kesavaraj and Dr. S. Sukumaran, "A Study on Classification Techniques in Data Mining", 4th ICCCNT - 2013, July 4 - 6, 2013, Tiruchengode, India.

[8] Sina Bahramirad, Aida Mustapha and Maryam Eshraghi, "Classification of Liver Disease Diagnosis: A Comparative Study", ISBN: 978-1-4673-5256-7/13, Pg 42 - 46, IEEE, 2013.

[9] McIver DK, Friedl MA. Using prior probabilities in decision-tree classification of remotely sensed data. Remote sensing of Environment. 2002 Aug 1;81(2-3):253-61.

[10] Sneha Chandra and Maneet Kaur, "Creation of an Adaptive Classifier to Enhance the Classification Accuracy of Existing Classification Algorithms in the Field of Medical Data Mining", International Conference on Computing for Sustainable Global Development, ISBN 987-93-80544-14-4, pg 1.188-1.193, IEEE, 2015.

[11] Nazanin Shahrokhi, Roxana Dehzad and Soheila Sahami, "Targeting Customers with Data Mining Techniques: Classification", International Conference on User Science and Engineering, ISBN: 978-1-4577-1655-3/11, Pg 212 - 215, IEEE, 2011.

[12] S. Udhaya Kumar, H. Hannah Inbarani and S. Senthil Kumar, "Bijective Soft Set based Classification of Medical Data", International Conference on Pattern Recognition, Informatics and Mobile Engineering, ISBN: 978-1-4673-5845-3/13/\$31.00, Pg 517 - 521, IEEE, 2013.