International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Efficient Multilevel Polarity Sentiment Classification Algorithm using Support Vector Machine and Fuzzy Logic

Vamshi Krishna. B, Ajeet Kumar Pandey, Siva Kumar A.P

Abstract*: This paper discusses an efficient algorithm for sentiment classification of online text reviews posted in social networking sites and blogs which are mostly in unstructured and ungrammatical in nature. Model proposed in this paper utilizes support vector machine supervised learning algorithm and fuzzy inference system for enhancing the degree of sentiment polarity of text reviews and providing multilevel polarity categories. Model is also able to predict degree of sentiment polarity of online reviews. The model accuracy is validated on twitter data set and compared with another earlier model.

Keywords: Opinion mining, Sentiment Classification, Machine learning, Fuzzy logic.

I. INTRODUCTION

As the usage of posting online opinions, recommendations, ratings, trending events or topics in public forums, social networking sites and blogs have gained more popularity in and become an excellent platform for exchanging opinions and ideas. Opinion mining and sentiment analysis are related in a sense that opining mining deals with analyzing and summarizing expressed opinions whereas sentiment analysis classifies opinionated text into positive and negative. Sentiment analysis or classification is an automatic method of predicting user emotions or sentiments which is hidden in the text reviews.

Twitter is a microblogging platform. The nature of the text reviews of twitter data source is in short form, with so many abbreviations and slang. Most of the opinions expressed in the form of text are mostly in unstructured format and ungrammatical in nature. Text may contain multiple emotions with different intensities. To handle discrete emotions there is need of multi-level sentiment classification techniques.

Many traditional machine learning algorithms are used in opining mining and sentiment analysis of text data and perform only binary classification either positive or negative. The sentiment polarity value of text reviews does not have a crispy value and is most fuzzy in nature.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
B. Vamshi Krishna*, Research Scholar, Dept. of CSE, JNTUA, Anantapuramu, India. vamshikrishna.bs@gmail.com
Dr. Ajeet Kumar Pandey, Delivery Manager, Dept. of RAMS, L&T Technology Services, Bangalore India. ajeet.mnnit@gmail.com
Dr. A.P Siva Kumar, Asst. Professor, Dept. of CSE, JNTUA, Anantapuramu, India. sivakumar.cse@jntua.ac.in

To increase the degree of sentiment polarity values of text reviews, fuzzy inference system is integrated with support vector machine classification model. This model gives flexibility to extend the degree of sentiment polarity for even cross domain sentiment classification.

II. RELATED WORKS

In recent years most of the researchers classified sentiment in the text using supervised learning techniques into positive, negative and neutral. Unsupervised techniques like topic models were used for predicting the trending topic and events.

Srishti and Susan applied fuzzy rule based unsupervised approach for computing sentiment classification of twitter data set with three class using multiple lexicons 1.

Erman et al. used adaptive neuro-fuzzy inference for predicting the emotional states in twitter time series data 2.

Hasan and Yuksel proposed a hybrid fuzzy decision-making approach to the investors risk profiles by using type- 2 fuzzy sets 3.

Sheeba and Vivekanandan proposed a framework for identifying implicit and explicit expression of emotions and topic detection by using fuzzy C-means clustering algorithm for grouping the words 4.

Vamshi et al. used fuzzy logic for the feature-based opinion mining of product reviews 5.

Subhashini proposed a frequent pattern mining and fuzzy logic-based opinion classification model to train a classifier based on concept similarity 6.

Pandey and Goyal 78 used fuzzy logic for early software fault prediction and improved reliability of software systems.

Prachi et al. proposed 3 - phase system for aspect-based sentiment summarization by using fuzzy logic 9.

Reddy et al. used expert model for identifying emotions in text which learns different set of features 10.

III. RESEARCH BACKGROUND

Fuzzy logic is integrated with machine learning technique for the sentiment classification purpose with various level of categories. fuzzy inference framework is designed which consists of fuzzification of input variables, Rule evaluation, aggregation of rule output and defuzzification.

The fuzzy inference framework is shown in the Fig. 1. Crisp values of sentiment and subjectivity obtained for all text reviews are given as input to fuzzifier and are converted to fuzzy values and rules are evaluated to predict the degree of polarity of reviews. Defuzzifer converts fuzzy value back to crisp output.

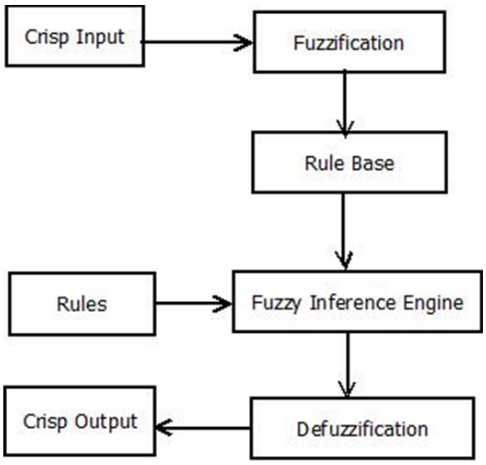

Fig. 1.Fuzzy Inference Framework

A. Support Vector Machine (S.V.M)

S.V.M is a supervised machine learning algorithm which can effectively classify the sentiment of textual reviews into positive and negative. S.V.M classify opinion words good, excellent as positive and bad, worst as negative reviews as shown in Fig. 2.

Fuzzy inferencing system is used for providing more levels of sentiment polarity of the text reviews. The opinionated words good, excellent and bad, worst will be given different level of polarity for effective sentiment classification and opinion summarization.

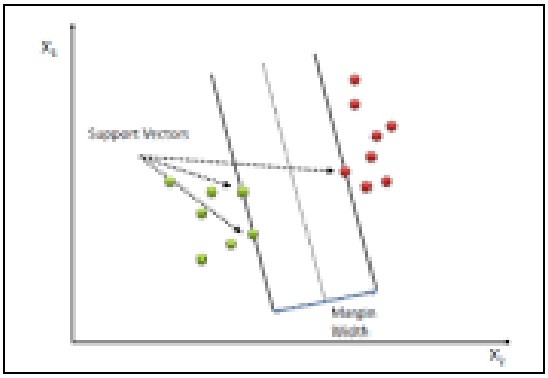

Fig. 2. SVM Model

B. Fuzzy Sets

Fuzzy sets are capable of having elements with partial degree membership. Polarity and subjectivity crisp values of all considered text reviews are computed and converted to fuzzy sets.

Let FP is a fuzzy set denoting polarity or subjectivity of review features, f1 , f2 ,f3. . . fn and $\mu_1, \mu_2, \mu_3 .. \mu_n$ are the membership value of features. By using Zadeh's notation 11, the degree of polarity of a feature can be represented using these membership values as:

$$F\acute{P}=\left\{\frac{\mu_1}{f_1}+\frac{\mu_2}{f_2}+\frac{\mu_3}{f_3}+..+\frac{\mu_n}{f_n}\right\}$$

C. Fuzzification

Fuzzification is the process of converting crisp or real values inputs into fuzzy set by using partial degree of membership. The individual values of polarity and subjectivity of the text reviews are input variables and are fuzzified by using triangular membership functions and output of the model which is degree of polarity. Table- I shows the division of five linguistic categories of input and output variables.

Table- I: Fuzzy Sets

Linguistic Categories	Ranges
Very Low (VL)	(0; 0; 0.2)
Low (L)	(0; 0.2; 0.4)
Medium (M)	(0.2; 0.4; 0.6)
High (H)	(0.4; 0.6; 0.8)
Very High (VH)	(0.6; 0.8; 1.0)

D. Fuzzy Rule System

The fuzzy rules are designed by using If-Then rules with linguistic variables as their antecedents and consequents to analyze text reviews are shown in Table-II.

Table-II: Fuzzy Rules

Polarity (Input variable)	Subjectivity (Input variable)	Degree of Polarity (Output variable)
Very Low (VL)	Very Low (VL)	Very Low (VL)
Low (L)	Low (L)	Low (L)
Medium (M)	Medium (M)	Medium (M)
High (H)	High (H)	High (H)
Very High (VH)	Very High (VH)	Very High (VH)

E. Defuzzication

The final step in fuzzy rule-based system is defuzzification. Output of the defuzzification is a crisp output, if corresponding fuzzy sets and membership functions are given. Centroid method is used as defuzzier to convert the fuzzy values of sentiment back to crisp value.

IV. PROPOSED WORK

The model architecture is shown in the Fig.3, which uses machine learning techniques and fuzzy logic for enhancing the degree of sentiment polarity of text reviews of twitter data sets. Support vector machine (S.V.M) is used for sentiment classification purpose which is implemented by using Python libraries Sklearn, and numpy packages. Skfuzzy package is used for building the fuzzy inference model.

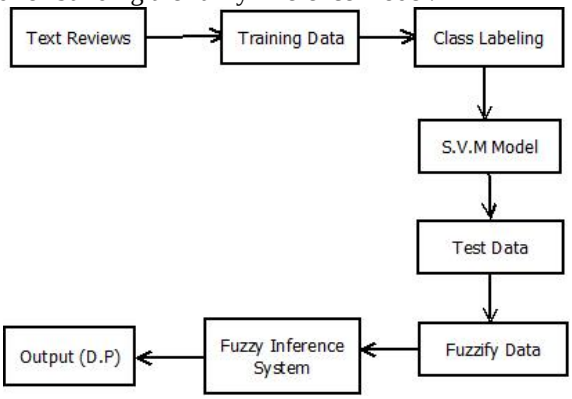

Fig. 3. Proposed Model

A. Data-set Used

Twitter streaming text data set is used in which recent reviews are used for opinion exchange related to various domains are collected. Data set is split into training, testing and validation.

B. Data-Preprocessing

Twitter reviews after filtering retweets are preprocessed by removing stop words, punctations, digits and other special symbols. These cleaned text reviews are split into tokens and converted into word vectors.

C. Proposed Algorithm

Below is the proposed algorithm:
Input: Text reviews
Output: Degree of Polarity (DP) of sentiment of the reviews

1) Identify the sentiment class label C {P, N} for the training data
   // C is a class label of the review which is either Positive (P) or negative (N)
2) Apply S.V.M model and fit model for target data
3) Classify the target data into classes say Positive (P) and Negative (N)
4) Develop fuzzy profiles with polarity and subjectivity of text reviews as input variables
5) Design fuzzy inference system (FIS) with rule evaluation to analyze text reviews
6) Output of the model is the Degree of polarity (DP)
7) Select the positive reviews and find its degree of positivity
8) Select the negative reviews and find its degree of negativity
9) Predict the overall sentiment of the target data on the basis of degree of polarity.

V. RESULTS AND DISCUSSION

A. Classifier Accuracy

S.V.M and Naïve Bayes (N.B) algorithms are applied on the data set and accuracy scores are compared using table III and table IV respectively and found S.V.M algorithm has higher accuracy scores. Also, confusion matrix which contains information about actual and predictions done by the classifier are shown in Fig. 4 and Fig. 5 respectively.

Table- III: S.V.M Classifier Accuracy

SVM Algorithm	Accuracy Score = 89.06			
	Precision	*Recall*	*F1score*	*Support*
Class 0	0.87	0.98	0.92	992
Class 1	0.95	0.71	0.82	508
Accuracy			0.89	1500
Macro avg	0.91	0.85	0.87	1500
Weigthed avg	0.9	0.89	0.89	1500

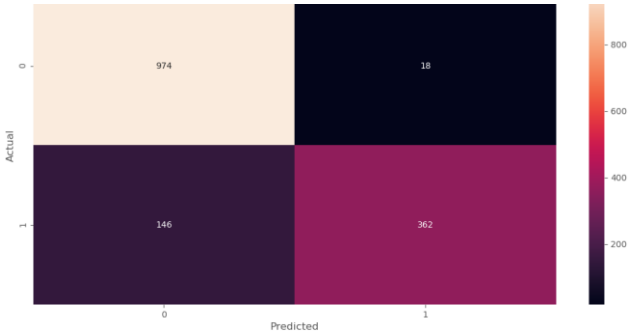

Fig. 4. S.V.M Confusion Matrix

Table- IV: Naïve Bayes (N.B) Classifier Accuracy

NB Algorithm	Accuracy Score = 83.2			
	Precision	*Recall*	*F1score*	*Support*
Class 0	0.81	0.97	0.88	982
Class 1	0.9	0.58	0.7	518
Accuracy			0.83	1500
Macro avg	0.86	0.77	0.79	1500
Weigthed avg	0.84	0.83	0.82	1500

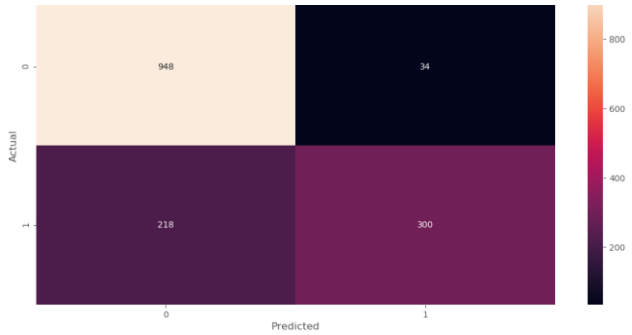

Fig. 5. N.B Confusion Matrix

B. Fuzzification results

Below are the fuzzy profiles for the input variables polarity and subjectivity of text reviews and output of the model is the degree of polarity of sentiment value.

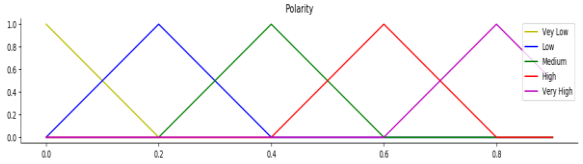

Fig. 6. Polarity

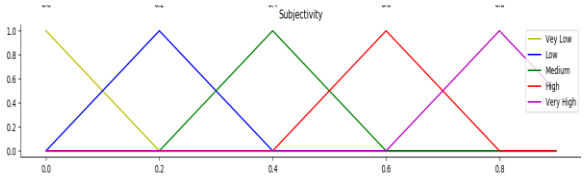

Fig. 7. Subjectivity

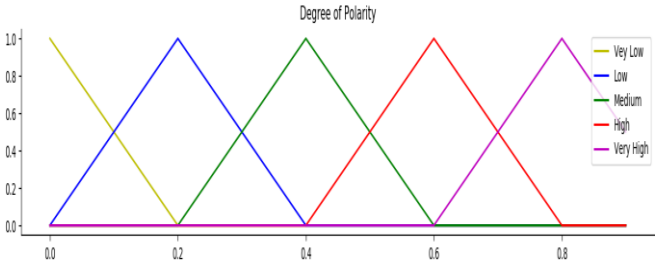

Fig. 8. Degree of Polarity

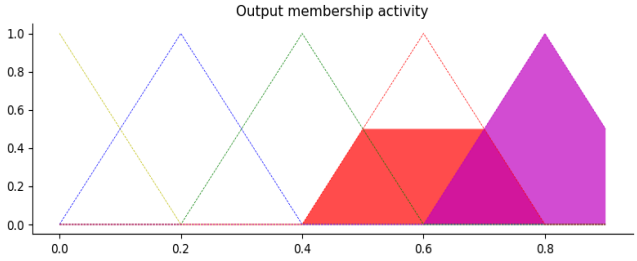

Fig. 9. Output membership activity

C. Defuzzification results

Centroid method is used as defuzzifier to get the crisp value back from the fuzzified data sets. Centroid values are shown in Table-V for few reviews of the test data set.

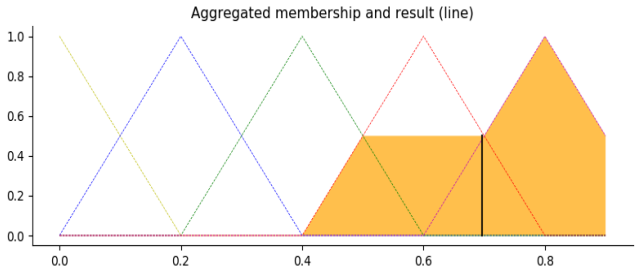

Fig. 10 . Aggregated membership

Table-V: Centroid Value of Fuzzy Profiles

Polarity	Subjectivity	Degree of Polarity
0.145833	0.595833	0.384
0.4	0.8	0.51046
0.207143	0.114286	0.18901
0.9	1	0.78095
0.55	0.285714	0.37763
0.7	0.6	0.54806

VI. CONCLUSION

A fuzzy logic-based model for opinion mining and sentiment analysis for the twitter data set has been presented by integrating with machine learning techniques. Model is also able to predict degree of sentiment polarity of online reviews. Model results are promising when compared with some of the earlier models for the twitter data sets. Some useful directions for future work include detecting opinion spam or fake review of textual reviews.

REFERENCES

1. Vashishtha S, Susan S. “Fuzzy rule based unsupervised sentiment analysis from social media posts,”. Expert Systems with Applications. 2019 Dec 30;138:112834.
2. Çakıt, Erman, Waldemar Karwowski, and Les Servi. "Application of soft computing techniques for estimating emotional states expressed in Twitter® time series data." Neural Computing and Applications: 1-14.
3. Dincer, Hasan, and Serhat Yuksel. "IT2-based fuzzy hybrid decision making approach to soft computing." IEEE Access 7 (2019): 15932-15944.
4. Sheeba, J. I., and K. Vivekanandan. "A fuzzy logic based on sentiment classification." International Journal of Data Mining & Knowledge Management Process 4.4 (2014): 27.
5. Krishna, B. Vamshi, Ajeet Kumar Pandey, and AP Siva Kumar. "Feature Based Opinion Mining and Sentiment Analysis Using Fuzzy Logic." Cognitive Science and Artificial Intelligence. Springer, Singapore, 2018. 79-89.
6. Subhashini LD, Li Y, Zhang J, Athukorale A. Opinion Classification using Pattern Mining and Fuzzy Logic. In2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer) 2018 Sep 26 (pp. 1-7). IEEE.
7. Pandey, A.K., and N.K. Goyal. 2009. “A Fuzzy Model for Early Software Fault Prediction Using Process Maturity and Software Metrics.” International Journal of Electronics Engineering 239–245.
8. Pandey, A.K., and N.K. Goyal. 2010. Predicting Fault-prone Software Module Using Data Mining Technique and Fuzzy Logic. International Conference [ICCT-2010], 3rd-5th December 2010.
9. Jain, Prachi, Anubhav Srivastava, Vineet Singh, and Bramah Hazela. "Aspect Based Sentiment Analysis by Fuzzy Logic." (2019).
10. Oota, Subba Reddy, et al. "Affect in Tweets Using Experts Model." arXiv preprint arXiv:1904.00762 (2019).
11. Zadeh, L.A. 1965. Fuzzy Sets. Information and Control 8: 338–353.

AUTHORS PROFILE

B. Vamshi Krishna received M.S degree in Software Engineering from B.I.T.S Pilani and pursuing Ph. D from JNTUA University. His research includes Data mining, Machine learning, Natural Language Processing and Information retrieval and extraction. He has publications in 1 national and 2 international conferences.

Dr. Ajeet Kumar did his M. Tech (Computer Science) and Ph.D. (Reliability Engineering) from MNNIT, Allahabad (India) and IIT Kharagpur (India) respectively. He is having 18 years of experience in System Assurance and RAMS. His broad is of research includes System Assurance, Software Safety and Reliability. He has to his credit, a number of research papers/article/books on System Assurance. Dr. Kumar is currently working as Delivery Manager – Railway at L&T Technology Services, Bangalore, India.

Dr. A. P Siva Kumar did his B. Tech from JNTUH, M. Tech from JNTUA, Ph.D from JNTUA in area of “Information Retrieval and Cross Lingual Intelligent Systems”in Year 2011. Recipient of Carrer Award for Young Teachers (CAYT) for the Financial Year 2013-14 with grant of Rs 1,50,000/ by AICTE, New Delhi. He is currently Assistant Professor in the Department of Computer Science and Engineering, JNTUA. Developed Examination Management Software “JEMS” JNTUA Examination Management System (EMS) which automates various tasks and procedures associated with the pre-examination and the post-examination phases associated with the Examination branch of an Autonomous College. Currently the Software is in live at JNTUCE Pulivendula and Audisankara College, Gudur. Master Trainer of Associate Analytics Trained by Nasscom in association with APSSDC.