

Performance Analysis of Feature Selection for Twitter Sentiment Analysis: Classification approach

Shweta V. Raut¹, Madhu M. Nashipudimath²

Student, Information Technology Department, Pillai College of Engineering, New Panvel, India¹

Assistant Professor, Department of Computer Engineering, Pillai College of Engineering, New Panvel, India²

Abstract: The social media platform is important as a means to spread information, opinions or awareness about real events. Twitter is prominent because of the huge amount of messages on all kinds of topics posted every day. Text classification of Twitter is never a trivial task that can be handled by common natural language processing techniques. There is also no agreement as to the definition of what type of operation will be performed in the event identification and classification of Tweets. Existing methods are difficult to reproduce and compare because they are often focused on certain types of events according to certain assumptions. This approach is in case of distinct nature. Previous work neglects impactful words due to the use of Bag-of-Words model. The contribution is for feature selection, Information Gain (IG) with consideration of adjectives as most impactful feature in POS tagger is used. Similarity Model, Relevance, Specificity & Span of relationship computation is also carried out later. SVM classification approach is implemented to construct positive and negative sentiments terms. Further this trained dataset is used for SVM classification of test dataset. Results show promising improvements.

Keywords: Feature Selection, Information Gain, Support Vector Machine, Sentiment Classification, Sentiment Analysis

I. INTRODUCTION

Sentiment analysis (SA) is the cognition of crowd's speculation, notions and feelings about a subject. Subjects can imitate individuals, incident or matters. This topic is likely to be covered in the comments. These two types of representational sentiment analysis can be used interchangeably. They expressed their mutual meaning. However, some researchers believe that concepts of viewpoint mining and SA are somewhat diverse [1], [2], [3]. Viewpoint mining excerpts and figures out the decisions of subjects, and SA pinpoints and diagnoses emotions intended in the content. Hence, SA goal is to investigate concerns, investigate desires explicated, and then segregate polarity into negative or positive.

Cyber content scrutiny is turning progressively significant due to communication boosting via. Web origins like; Blogs, online mail, forum sites, wikis and societal networks. In the process of SA, it is procedure to determine and diagnose decisions and beliefs, influences and bigotry. Historic scheme of testimony diagnosis, like, topic diagnosis, may not be potent on today's web testimony data. In divergence, cyber discourse is bloated in information associated to emotions [6], [7]. As a result, SA has recently been practiced with several schemes of cyber-based discourse [8]. Applying to zealot crowd forums gives you intuition into significant discussions and tendencies.

II. BACKGROUND STUDY

In present technological globe person comments the everyday intensions over cyberspace. With the launch of societal structure mediums, organizations have initiated to utilize societal as yielding land for sentimental mindset. But, the opposition is to deal with abundance decisions into glided knowledge. SA estimates catalog the decisions by assorting comments in binary divisions polarity.

Amir Hossein Yazdavar, Monireh Ebrahimi and Naomie Salim [21] developed a Frizzy based cognition constructing model for emotion categorization of extraordinary group of judgements that includes diversity from desiderate extent or limit. The main conclusion is to enhance the correctness of recent drug viewpoint mining systems where dataset is collected from www.askapatient.com. Corpus which contains various type of quantitative medical term can be enlarged and also more medical expert annotator can be consulted in future to erect a superior gold classic corpus and obtain enhanced medical cognition.

Geetika Gautam and Divakar Yadav [22] performed SA on Product reviews for highly unstructured data. Naive Bayes, Maximum Entropy and Support Vector Machine are the sentiment classifiers used. The crucial intent is to analyze an abundant measure of audits by using twitter dataset which are already labelled. Naive Bayes method is superior as compared to Maximum Entropy and SVM is superior alone.

Lei Zhang, Rdhiman Ghosh, Mohamed Dekhil, Meichun Hsu and Bing Liu [23] invented a naive subject-level emotions scrutiny with immense precision at poor recall. Supplementary tweets are those which are likely to be optionated are recognized in automated manner to enhance recall. Further manipulating facts in the outcome of the lexicon-based technique Chi Square test is carried out for feature selection.

Shahana P.H., Bini Omman [24] used 2000 audit data belonging to Bollywood/Hollywood films. The methods like Mutual Information, Chi Square, Information Gain and Tf-idf are used to choose characteristics from immense extent of characteristic set. Using Information Gain for unigram with stemming and with stop words outcomes poor result as compared to that with unigram with stemming and without stop words for both classes.

Ankit Pradeep Patel, Ankit Vithalbhair Patel, Sanjaykumar Ghanshyambhai Butani and Prashant B. Sawant [25] discussed an approach to publish stream of tweets on electronic products. The accurate result is obtained using SVM when compared with supervised algorithm like Support Vector Machine (SVM). The accurate result is obtained using SVM when compared with Vector Quantization method.

Munir Ahmad, Shabib Aftab, Muhammad Salman Bashir, Noureen Hameed [26] focused on analytical scrutiny of literature from year 2012 to 2017 on SA. Support Vector Machine (SVM) is compared with other approaches. The datasets used are movie and software reviews. In future other techniques can be used for comparative analysis using same datasets.

Prachi Bansal and Ramanjot kaur [27] introduced a hybrid approach of using Swarm Intelligence optimization algorithms with Support Vector Machine (SVM) and Naive Bayes (NB) classifiers. These method solves the problem of dual optimization. Datasets used are collected from various websites like Stand ford University. The result obtained here is that SVM doesn't works well as compared to that of Naive Bayes algorithm. The accuracy of SVM can be increased when used with other optimization methods.

Till present, surveyors have deal with many methodologies to provide answers to the issues.

Table 2.1 Summary of Literature Survey

| Sr. No | Title of the paper, Author & Year of Publication | Observations/Remarks |
|--------|---|--|
| 1 | On the Feature selection and Classification based on Information Gain for Document Sentiment Analysis, Asriyanti Indah Pratiwi and Adiwijaya, 2018 [28] | To find an optimum features selection and classification. Information Gain calculation (IGC) outperforms out of all other classifiers. SVM and ANN encountered over fitting issue. |
| 2 | Twitter Sentiment analysis of Game reviews using machine learning techniques, T.D.V. Kiran, K. Gowtham Reddy, Jagadessh Gopal, 2017 [29] | To obtain a highly accurate model of sentiment analysis of tweets. Maximum Entropy outperforms SVM and NB. |
| 3 | The effects of Pre-processing strategies in Sentiment analysis of online movie reviews, Harnani Mat Zin, Norwati Mustapha, Masrah Azrifah Azmi Murad and Nurfadhlina Mohd Sharef, 2017 [30] | To probe the reaction of Pre-processing policies in SA.SVM with linear kernel doesn't performs well as compared to that of SVM non-linear kernel for both TF and TF-IDF. |
| 4 | Sentiment Analysis of Tweets using SVM, Munir Ahmad, Shabib Aftab and Iftikhar Ali, 2017 [31] | Analyzed two datasets of tweets using SVM. 1) Self driving cars, 2) Apple products. SVM works well for Apple products datasets. |
| 5 | An Information Gain-Driven Feature study for Aspect-Based Sentiment Analysis, Kim Schouten, Flavius Fransincar and Rommert Dekkar, 2016 [32] | It analyzed which characteristics are superior predictors for Feature-level viewpoint. |

| | | |
|---|---|--|
| 6 | A Survey on Sentiment Analysis of (Product) Reviews, A. Nisha Jebaseeli and E. Kirubakaran, 2012 [33] | Paper discussed an audit on SA emotion mining related to commodity audits. In emotion mining, certaining emotions leads to be effortless but certaining subjects and their equivalent characteristics is tough. Fusing both the task is very tough and also correctness is the issue. |
| 7 | Sentiment Analysis in Twitter with Lightweight Discourse Analysis, Subhabrata Mukherjee, Pushpak Bhattacharyya, 2012 [34] | Proposed an inconsequential technique for utilizing discourse relations for categorization discovery of tweets. In the absenteeism of parsing and tagging information, due to the boisterous nature of the tweets, the extent of the discourse marker has been curiously taken till the sentiment border or till the next discourse marker. Does not contemplate the positional significance of a discourse marker in the sentiment but contemplate all the discourse markers to be equitably significant. |

III. FRAMEWORK

We have presented a framework in which we have explained a process of pre-processing, feature selection, various other computations and classification. Twitter dataset is provided as input to Pre-processing stage that gives output as cleaned dataset with removal of stop words, hyperlinks, special characters, text converted from uppercase to lowercase to treat all the words equal i.e., Happy=happy. The datasets being used for the evaluation of the proposed Sentiment analysis based on Relevant Feature Selection and SVM Classification Approach is being retrieved from "www.kaggle.com". Kaggle provides Open Source Datasets in multiple domains and various kind of twitter data collection data.

For exclusive impression of a word, a set of synonymic words are recorded in Sentiment Word Dictionary (SWD). Aimless selection is done for nouns from SWD and excerpt a pair of synonymic words from a dictionary of every selected term. If a selected noun is cryptic, then acknowledgement of the dictionary for the assertive sentiment senses is done, thus, in such case adjectives are considered to differentiate the cryptic words. In Feature extraction stage, Information gain (IG) is carried out along with Similarity model and Relevance, Specificity & Span of relationship calculations by fetching for synonymic and antonymic words from SWD and by referring the trained labelled data formed during training the data for further SVM classification.

The morphological accord among semantic approaches are customarily associated depends upon three criteria including exactness, recall and correctness. These criteria are computed for the sentiment concepts can be related to each other by referring to multiple twitter data related to the user application, categorized depends upon their location in the concept and the accords intending the semantic guild brings up the strength of linkage among concepts.

Based on the processing and measuring of the above methodology we predict the twitter datasets sentiment and perform the following evaluation measure.

The proposed system consists of:

3.1 Pre-processing

3.2 Feature Extraction using Information Gain Measure - Similarity model and Relevance, Specificity & Span of relationship calculations.

3.3 Sentiment Segregation

3.4 Sentiment Classification using SVM

The details of modules are as follows:

3.1 Pre-Processing:

The factual-world data stores are supremely vulnerable to noisy, absent terms, and differing data because of their representative supreme measurement and their probable dawn from multiple miscellaneous places. Moderate-grade data will provide moderate-grade outcomes. There are number of facts initial-processing methodologies. Facts purification can be applied to eradicate noise and do precise discrepancies in the fact. Fact conversions like; normalization can upgrade the correctness and efficiency of mining methodologies considering space dimensions. This module is responsible for eliminating Stop Words and Noisy data from the collected datasets.

This module involves in capturing sentiment features from real-time data from Twitter and storing it in a suitable format in support of Sentiment Word Dictionary (SWD) [35]. Based on the set of defined sentiments we generate a sentiment feature pattern. It will implement a correlation association mining approach to find the sentiment feature relation with the define SWD phrases to provide the sentiment feature selection.

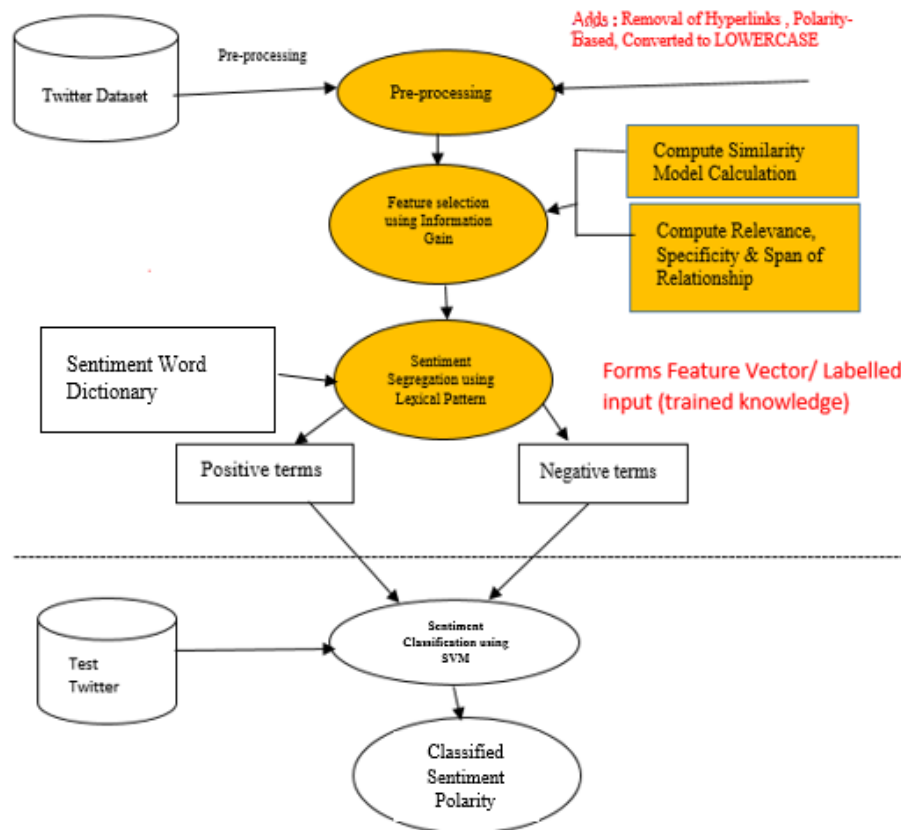


Figure 1 Proposed Architecture

3.2 Feature Extraction using Information Gain Measure (IG)

The processed dataset after pre-processing has a lot of distinctive properties. The feature extraction method extracts the aspect (adjective) from the dataset using information gain measure. Information gain is an attribute selection measure based on the value or “information content” of messages. Let assume a sentence N represent or hold the terms adjective of phrase D. The adjective with the highest information gain is chosen as the splitting adjective for sentence N. This adjective minimizes the information needed to classify the terms in the resulting phrase and reflects the least randomness or “impurity” in this adjective.

After pre-processing stage, entropy is calculated for trained dataset & 10 most impactful features are being extracted using IG. Then, entropy is calculated for test dataset, and outputs of most impactful features are 3 or less terms, those can be seen at final stage. Thus, while training data it takes time but while applying different dataset to test it requires less time as only 3 features to fetch from trained dataset as positives & negatives. The output after computation of test data is being added to train dataset. Similarity and Relevance, Specificity & Span of Relationships computations are computed, i.e., features polarities are fetched from trained dataset.

3.3 Sentiment Segregation:

This module implements the function to categorize the datasets and to provide an easy access to relevant words in terms of positive and negative terms. To do so it takes a support of Sentiment Word Dictionary. To segregate the sentiment from the extracted features we implement a lexical Pattern algorithm in compared to terms of Sentiment Word Dictionary.

Generally, a synonymic linkage between terms can be conveyed utilizing more than account of two divergent schemes, E is a F, and E is a large F. Both these schemes stipulate that there subsists an in-a linkage among E and F. The extracted features are then given as input to SVM categorizer for categorization where, you can actually understand the extraction of 3 features of test dataset and its category. [4.4-Step 5]

3.4 Sentiment Classification using SVM:

This module implements an SVM based classification process for semantic prediction from the datasets. To train SVM and to discern positive and negative term matches, we took an advantage of a training dataset term-matches obtained from feature extraction and sentiment segregation module. It is composed of synonymic term-matches of positive training samples and non-synonymic term-matches as a negative training samples. For each term-match we produced measurement feature vector of a term-match and finally, we train the SVM using labelled feature vectors.

After training SVM, positive training samples are term-matched, we will utilize it to access the semantic uniqueness between two given terms [36] ,[37]. Minimizing the similar techniques, we utilized to produce feature vectors for training, we fabricate the test datasets measurements of feature vector for a match of terms and among which we measure semantic uniqueness to classify the sentiments in the datasets.

- **SVM Improvisation for Semantic Classification**

The Support Vector Machine (SVM) classification generally learns for individual characteristic. Nevertheless, solely acquires knowledge of classifiers by fixing them to all characteristics semantic of the features precisely [38].

For each feature, we requisite to choose the characteristics that are frequently efficient in designing this feature. It is necessary to regulate this choice based on the two inspections:

Positive-level characteristics like these are extricated by IG feature selection that leads this, extrication may not focus to outline the particular feature and involve non-repeating knowledge. Thus, we requisite to choose illustrative and characteristics that are benefited to relate present semantic features. The exercise of electing a subspace of most pertinent characteristics has been acting a significant player in boosting up the learning stage and alleviating the consequence of the ambiguity of the term. We will operate feature election technique to select the most effective terms as positive and negative terms. In particular, if we want to learn a statement “I’m eating an apple while playing computer game” categorizer, we elect characteristics that outcomes well at differing the illustrations of data with positive terms “apples” and with negative term “apples”. By following such methodology, we benefit baffled about “apple” or “computer”, as both types are related to “fruits” or have “computer” depends on their positivity or negativity similarity.

Consider, two terms R and S, we fix the issue of aligning similarity among R and S, as one of erecting a mathematical component $\text{sim}(R,S)$ that outcomes a computation in the length[0,1]. If R and S are eminently identical to the sentiment dictionary, it expects $\text{sim}(R,S)$ to be near to 1. On the next side, if R and S are not statement-wise identical, then we expect $\text{sim}(R,S)$ to be near to 0.

It constructed plentiful characteristics that outcomes the uniqueness among R and S using terms computes and statements recovered from input terms. Utilizing such characteristic depiction of terms, we train a binary-class SVM to categorize the positive and negative sentiment.

- **Semantic Classification Measures:**

Firstly, by measuring Exactness, Recall and Correctness rates to validate the working of the proposed system.

Secondly, by utilizing training data to acquire a categorizer or predictor and then to project the correctness of the out-coming learned system can outcome in evasive over-idealistic acquisition due to over-pursuable of the learning technique to the data. Rather, correctness outperforms in measurement on a test dataset including class-labelled rows that were not utilized to instruct the system. The correctness of a categorizer on an available test set is the computation of percentile of the categorizer. In the pattern realization survey, this is described to as the overall recognition rate of the categorizer, that is, it gives rise to have correctly categorizer to recognize tupelos of several classes.

Table 3.1 - A Confusion matrix for positive and negative

| | | Predicted class | |
|----|--|-----------------|-----------------|
| | | C1 | C2 |
| C1 | | true positives | false negatives |
| C2 | | false positives | true negatives |

Actual class: SA process further to utilize relevant data as training data for problematic domains. Also, the work can be extended to consider sarcastic reviews. Twitter datasets used in the work is collection of data from www.Kaggle.com site. Work is carried out in JAVA platform using Eclipse IDE. Information Gain (IG) and Support Vector Machine

(SVM) are used for feature selection and classification respectively. The result shows the accuracy in precision and accuracy. A Confusion matrix is a beneficial tool for two classes as shown in Figure-3.1. Consider p classes, a Confusion matrix is a relation of at least dimension p by p . An entry

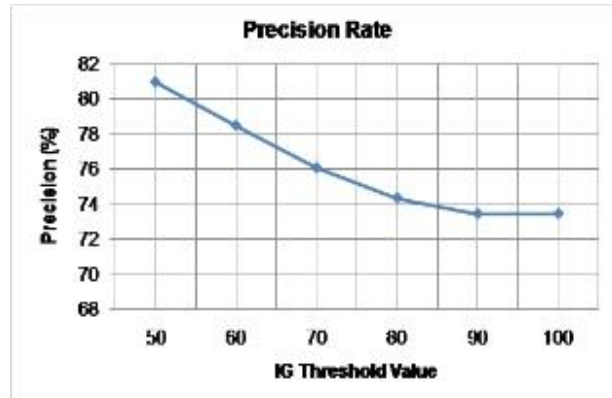
$CM_{i,j}$ is the first p rows and p columns involves the number of tuples of class i that were segregated by the categorizer as class j . For a categorizer to have precise correctness, ideally most of the tuples would be arranged including the tangent of the Confusion matrix, from entry $CM_{i,i}$ to entry $CM_{j,j}$ with the rest of the entries being near to zero. The tabular representation depicts tabular rows and columns to present total rates per class.

Consider two classes, i.e., positive against negative tuples. True positives refer to positive tuples that were rightly labelled by the categorizer. False positives are the negative tuples that were non-precisely labelled. Similarly, false negatives are the positive tuples that were non-precisely labeled.

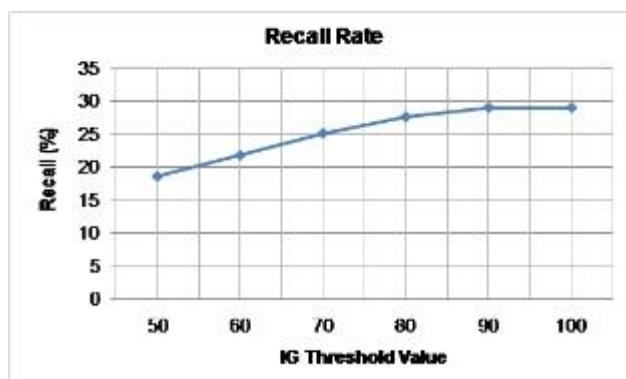
In this paper, firstly pre-processing process is carried out that includes removal of Stop-words, Special Characters, Stem-words, Hyperlinks. Uppercase terms are converted to LOWERCASE. Then, the remaining terms are considered as input to Information Gain module where most impactful terms are only considered. After applying, information gain algorithm, Similarity model calculation, Relevance, Specificity and Span of Relationships are computed for ambiguous terms comparison with SentiWordNet Dictionary synonymic and antonomic terms. During applying test dataset, the terms are matched with the result of trained dataset for categorization.

The result obtained is correct precision and accuracy. True positive result is obtained when Confusion Matrix is used for validation.

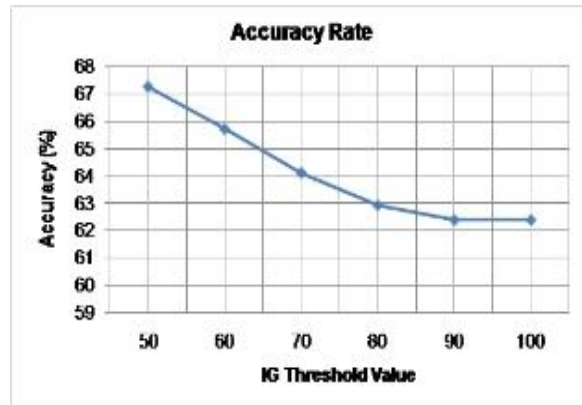
The context and client's decisions of the text is important to be considered in varietal applications. That is why it is very much needed to do deep survey on event-based Sentiment Analysis. In future, by using the emotion detection technology and NLP tools can be used to enhance



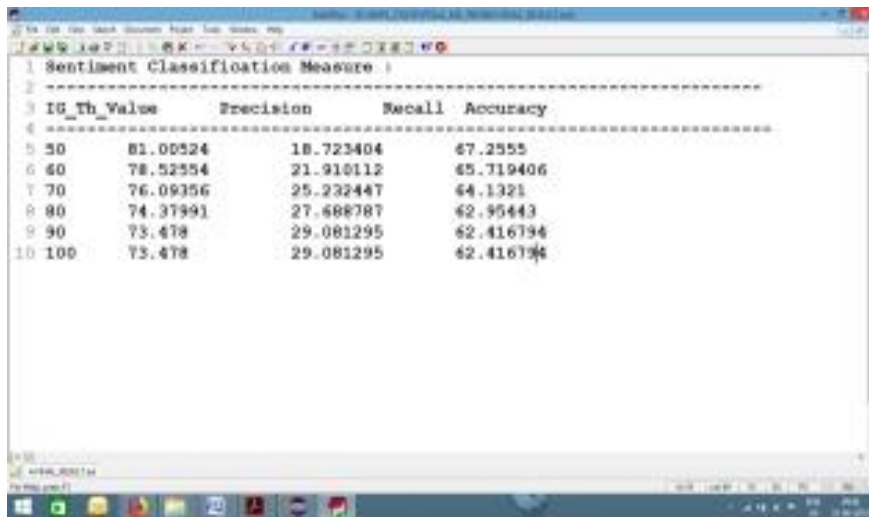
(a) Percentage of Precision Measure Vs Information Gain Threshold value



(b) Percentage of Recall Measure Vs Information Gain Threshold value



(c) Percentage of Accuracy Measure Vs Information Gain Threshold value



| IG_Th_Value | Precision | Recall | Accuracy |
|-------------|-----------|-----------|-----------|
| 50 | 81.00524 | 18.723404 | 67.2555 |
| 60 | 78.52554 | 21.910112 | 65.719406 |
| 70 | 76.09356 | 25.232447 | 64.1321 |
| 80 | 74.37991 | 27.688787 | 62.95443 |
| 90 | 73.478 | 29.081295 | 62.416794 |
| 100 | 73.478 | 29.081295 | 62.416794 |

(d) Sentiment Classification Measures Vs different Information Gain Threshold

CONCLUSION

From the above figures, at Information Gain Threshold value 50, Precision of 81% and Accuracy of 67.25% are obtained respectively. An experimental evaluation using Twitter Sentiment Dataset shows an improvisation in precision and accuracy measure. Recall is found to be increasing with the increase in IG threshold value. In the work, document-level polarity classification is attempted to categorize emotions in Bollywood or Hollywood reviews, public broadcast details, or societal forum reviews. Most impactful terms are considered using Information Gain (IG) followed by Similarity model calculation instead of using Bag-of-words for many domains terms if occurred. Figures (a), (b), (c) shows the overall working of the system. At Information Gain Threshold value 50, Precision of 81% and Accuracy of 67.25% are obtained respectively. An experimental evaluation using Twitter Sentiment Dataset (d) shows an improvisation in precision and accuracy measure. Recall is found to be increasing with the increase in IG threshold value.

ACKNOWLEDGMENT

It is a great pleasure and moment of immense satisfaction for me express my profound gratitude to my dissertation Project Guide, **Prof. Madhu Nashipudimath** whose constant encouragement enabled me to work enthusiastically. I would like to express my special gratitude to ME. Co-ordinator and HOD of Information Technology **Dr. Sharvari Govilkar** who are always been there to help and give her time and advice in spite of her busy schedule. I am also thankful to **Dr. Sandeep M. Joshi, Principal**, Pillai College of Engineering, New Panvel, for his encouragement and for providing an outstanding academic environment, also for providing the adequate facilities. My deep sense of gratitude to IJIRCST for giving me an opportunity for publishing my manuscript.

REFERENCES

- [1]. S. Joscha, H. Feuerriegel, Prendinger, "Improving Sentiment Analysis with Document-Level Semantic Relationships from Rhetoric Discourse Structures", Proceedings of the 50th Hawaii International Conference on System Sciences, ISBN:978-0-9981331-0-2,2017.
- [2]. Aisopos, Fotis, "Content vs. Context for sentiment analysis: a comparative analysis over micro blogs." Proceedings of the 23rd ACM conference on Hypertext and social media. ACM, 2012.
- [3]. Ortigosa, Alvaro, J. M. Martin, and Rosa M. Carro, "Sentiment analysis in Facebook and its application to e-learning." Computers in Human Behavior 31:527-541,2014.
- [4]. M. A. Hearst, "Direction-Based text interpretation as an information access refinement", In Text-Based Intelligent Systems: Current Research and Practices in Information Extraction and Retrieval, 1992.
- [5]. E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," IEEE Trans. Antennas Propagat., to be published.
- [6]. P. Subasic, A. Huettner, "Affect analysis of text using fuzzy semantic typing", IEEE Transaction on Fuzzy Systems 9, 4, 483-496,2001.
- [7]. A. Alexandra, H. Jesu's M, Montoyo Andre's, "Detecting implicit expressions of emotion in the text: a comparative analysis", Decision Support System, Vol. 53, pp. 742-53, 2012.
- [8]. Agarwal, Apoorv, et al., "Sentiment analysis of Twitter data", Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, 2011.
- [9]. A. Nourbakhsh, et. Al., Newsworthy Rumor Events: A Case Study of Twitter. 2015 IEEE International Conference on Data Mining Workshop (ICDMW), IEEE Computer Society, Los Alamitos, CA, USA, v.00, n. undefined, p.27-32, 2015.
- [10]. M. Isa, V. Piek, "A lexicon model for deep sentiment analysis and opinion mining applications", Decision Support System Vol. 53, pp. 680-8, 2012.
- [11]. A. O'Connor, R. Balasubramavan, B.R Routledge, Noah A. Smith, "From tweets to polls: Linking text sentiment to public opinion time series", Proceedings of the International AAAI Conference on Weblogs and Social Media, Washington DC, May 2010.
- [12]. A. Hogenboom, F. Frasinca, F. De Jong and U. Kaymak, "Using Rhetorical Structures in Sentiment Analysis", Communications of the ACM, vol. 58, no. 7, pp. 69-77, 2015.
- [13]. H. Takamura, T. Inui and M. Okumura, "Extracting semantic orientations of phrases from the dictionary", Proceedings of the Joint Human Language Technology/ North American Chapter of the ACL Conference (HLT-NAACL), 2007.
- [14]. P. Chesley, B. Vincent, L. Xu, and R. Srihari, "Using verbs and adjectives to automatically classify blog sentiment", in AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW), pp. 27-29, 2006.
- [15]. W. Medhat, A. Hassan, H. Korashy, "Sentiment analysis algorithms and applications: A survey", Elsevier on behalf of Ain Shams University, 2090-4479, 2014
- [16]. M. Rodrigo, Valiati Joa~o Francisco, Gavia~o Neto Wilson P., "Document-level sentiment classification: an empirical comparison between SVM and ANN", Expert Syst. Application Vol. 40, pp. 621-33, 2013.
- [17]. M. Hu and Bing Liu, "Mining opinion features in customer reviews", In Proceedings of AAAI, pages 755-760, 2004.
- [18]. Adnan, Song Fei, "Feature selection for sentiment analysis based on content and syntax models", Decision Support System Vol. 53, pp. 704-11, 2012.
- [19]. R. Nilson, et al., "Evaluating Feature Selection for SVMs in High Dimensions", In Machine Learning: ECML 2006 SE-72, Springer, Comp. Science, v. 4212, p. 719-726, 2006.
- [20]. Bepaloy, B. Bai, A. Shokoufandeh, and Y. Qi, "Sentiment Classification Based on Supervised Latent n-gram Analysis", pp. 375-382, 2011.
- [21]. A. Hossein Yazdavar, M. Ebrahimi, N. Salim, "Fuzzy-Based Implicit Sentiment Analysis on Quantitative Sentences", Journal of Soft Computing and Decision Support Systems 3:4, 7-18, 2016.
- [22]. G. Gautam, Divakar Yadav, "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis", IEEE Conference, pp. 136-138, 2014.
- [23]. L. Zhang, R. Ghosh, M. Hsu, B. Lui, "Combining lexicon-based and learning-based methods for Twitter sentiment analysis", Hewlett-Packard Laboratories, HPL-20 11-89, 2011.
- [24]. Shahana P.H. And Bini Omman, "Evaluation of Features on Sentiment Analysis", Science Direct, Elsevier, Procedia Computer Science, 2015.
- [25]. Ankit Pradeep Patel, Ankit Vithalbhai Patel, Sanjaykumar Ghanshyambhai Butani and Prashant B. Sawant, "Literature Survey on Sentiment Analysis of Twitter data using Machine Learning Approaches", International Journal for Innovative Research in Science & Technology, 2017.
- [26]. Munir Ahmad, Shabib Aftab, Muhammad Salman Bashir and Noureen Hameed, "Sentiment Analysis using SVM: A Systematic Literature Review", International Journal of Advanced Computer Science and Applications, 2018.
- [27]. Prachi Bansal and Ramanjot Kaur, "Twitter Sentiment Analysis using Machine Learning and Optimization Techniques", International Journal of Computer Applications, 2018.
- [28]. Asriyanti Indah Pratiwi and Adiwijaya, "On the Feature selection and classification based on Information Gain for Document Sentiment Analysis", Applied Computational Intelligence and Soft Computing, 2018.
- [29]. T.D.V. Kiran, K. Gowtham Reddy and Jagadeesh Gopal, "Twitter Sentiment Analysis of game reviews using Machine Learning Techniques", Journal of Chemical and Pharmaceutical Sciences, 2017.
- [30]. Harnani Mat Zin, Norwati Mustafa, masrah Azrifah Azmi Murad and Nurfadhline Mohd Sharef, "The effects of pre-processing strategies in Sentiment Analysis of online movie reviews", AIP Conference Proceedings, 2017.
- [31]. Munir Ahmad, Shabib Aftab and Iftikhar Ali, "Sentiment Analysis of Tweets using SVM", International Journal of Computer Applications, 2017.
- [32]. Kim Schouten, Flavius Fransincar and Rommert Dekker, "An Information Gain-driven Feature study for Aspect-Based Sentiment Analysis", Semantic Scholar, 2016.
- [33]. A. Jebaseeli, Nisha, and E. Kirubakaran, "A Survey on Sentiment Analysis of (Product) Reviews" , International Journal of Computer Applications, 47.11, 2012.
- [34]. Subhabrata Mukherjee and Pushpak Bhattacharya, "Sentiment Analysis in Twitter with Lightweight Discourse Analysis", IBM India research lab, 2012.
- [35]. SentiWordNet Available online: <http://sentiwordnet.isti.cnr.it/> (accessed on 26 February 2018).
- [36]. Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirband, "Machine Learning-based Sentiment Analysis for Twitter accounts", Journal of MDPI, www.Mdpi.com/journal/MCA, Math. Computer. Applied. 2014.
- [37]. Sneha S. Mulatkar, "Sentiment Analysis: A Survey", International Journal of Innovative Research in Computer Science & Technology", ISSN: 2347-5552, Volume-2, Issue-2, March 2014.