Performance Analysis of Classification Methods for Opinion Mining

Sumathi T
Department of Master of Computer Application
Institute of Road and Transport Technology, Erode, Tamilnadu, India

Karthik S
Department of Computer Science and Engineering
SNS College of Technology, Coimbatore, Tamilnadu, India

Marikannan M
Department of Computer Science and Engineering
Institute of Road and Transport Technology, Erode, Tamilnadu, India

Abstract- Opinion mining extracts opinions expressed by a specified target from a set of documents. Opinion mining determines sentiment orientation positive, negative, or neutral of a (short) text and is a text understanding technology assisting people locate relevant opinions in a review collection of large volumes, quickly. A search engine based on opinion mining technology reveals potential to address the above dilemma. An opinion-mining tool extracts opinion units and saves them in an opinion database. The collection of ordered information through the opinion search engine is considered opinion based rather than document based one. Hence, product review information is easily accessed. Recently, opinion mining has received many attentions due to many forums, blogs, e-commerce web sites, news reports and additional web sources where people express opinions. Some opinion mining tasks try to classify detected opinion using various scales. The main objective of this method is opinion identification in a text and then to classify them as positive, negative or neutral. In this paper, it is proposed to extract the feature set based on IDF (Inverse Document Frequency) and classification models based on Naive Bayes, RIDOR and FURIA.

Keywords – Opinion mining, IMDb, Inverse Document Frequency (IDF), Naïve Bayes, Fuzzy Unordered Rule Induction Algorithm (FURIA), Ripple Down Rule Learner (RIDOR)

I. INTRODUCTION

Opinion analysis is an interesting research topic in both information extraction and knowledge discovery. Discovered opinions can be used in many applications. For example, product opinions help customer make purchase decisions while opinions, on different source related specific policies, improve government management. Also, as a natural language processing technology, opinion analysis promotes research in information extraction and knowledge discovery like automatic summarization [1].

Opinion mining is a research area in Text Mining (TM) designated by different terms like subjectivity analysis, sentiment analysis or sentiment orientation. There are many definitions for each. For example, Subjectivity Analysis is the recognition of opinion-oriented language to differentiate it from objective language. Sentiment Analysis classifies reviews based on their polarity (positive or negative). So, all terms refer to similar fields of study. Some opinion mining tasks aim to classify detected opinion through the use of varied scales. In many cases, the aim is identifying opinions in a text and classifying them as positive, negative or neutral. In other occasions, the goal assigning different rates like ‘‘very bad’’, ‘‘bad’’, ‘‘satisfactory’’, ‘‘good’’, ‘‘very good’’, or ‘‘excellent’’. Sentiments are ranked in a range of 1to 5 stars. Other systems use ‘‘thumb up’’ or ‘‘thumb down’’ notation [2].

Opinion mining attracts interest both in academia and industry due to its potential applicability. A promising application is an analysis of social networks opinions. Many write their opinions in forums, micro blogging or review websites. This data is useful for businesses, governments, and individuals, who track attitudes and feelings in such sites. Much data containing useful information is available, for automatic analysis. For instance, a customer who proposes to buy a product searches the Web to find other customers/reviewers opinions about the product. Such reviews affect the customer’s decisions. Opinion mining is the computational study of text expressed opinions, sentiments and emotions. Opinions for any entity/object/product/person exist on the Web and also for features/components of objects like, cell phone batteries, keyboards, touch screen displays, etc [3].

Opinion mining, data mining and computational linguistics sub discipline refers to computational techniques to extract, classify, understand, and assess opinions expressed in online news sources, social media comments, and user-generated content. Sentiment analysis uses opinion mining to identify sentiment, effect, subjectivity and emotional states in online text. For example, we could attempt to answer such questions:

• What were young US voter's opinions toward Democratic and Republican presidential candidates during the recent election?
• After September 11, how do international Jihadi forums introduce radical ideology/incite young members?
• What are investors, employees, and activists opinions/comments toward Wal-Mart in view of its cost-reduction moves and global business practices? [4].

IMDb[5] an Internet Movie Database, is a collection of movie information (claimed to be earth's largest movie database) which started as a hobby by an international movie fans groups and currently belongs to Amazon.com. IMDb catalogs all details about a movie, from who was in it, to who made it, and related trivia, to filming locations, and even where reviews and fan sites can be located on the web. The IMDb web site (http://www.imdb.com/) ensures 49 text files in ad-hoc format (lists) with varying characteristics about movies (actors. list or running-times. list). Lists content is continually updated/enlarged; as of October 5th 2006, movie lists had more than 858.000 movie titles.

Movie Lens is a movie recommender project, developed by the University of Minnesota's Department of Computer Science and Engineering is a collaborative filtering system collecting movie preferences from users and users groups having similar tastes. Based on user/groups expressed movie ratings it attempts to predict for each individual's opinion on movies they have yet to see. The Movie Lens web site (http://movielens.umn.edu/) has two data sets available. The first includes 100,000 ratings for 1682 movies by 943 users and the second consists of approximately 1 million ratings for 3883 movies by 6040 users [6].

Text Mining is the discovery of new and previously unknown information, through automatic information extraction from various written resources. The main element is linking extracted information together to form new facts/new hypotheses for further exploration through conventional experimentation. Text mining is different from web search in which a user searches for something already known and written about by someone else.
Text mining is a variation of data mining that finds interesting patterns from large databases. The other names for Text mining are Knowledge-Discovery in Text (KDT), Text Data Mining, or Intelligent Text Analysis is the process of extracting interesting/non-trivial information/knowledge from unstructured text. Text mining is an interdisciplinary field drawing on information retrieval, data mining, machine learning, statistics and computational linguistics [7].

The RIpple DOwn Rule Learner (RIDOR) algorithm learns rules with large support, but then qualifies these by also learning exceptions to each rule. The exceptions with the maximum support are chosen, and then exceptions are sought for these exceptions, and so on, in an iterative manner.

This work proposes to investigate FURIA and RIDOR classification efficiency using a movie review data set. Section 2 reviews some work in literature. Section 3 describes methodology used with sections 4 and 5 showcasing results and discussion.

II. LITERATURE SURVEY

An unsupervised learning algorithm was proposed by Turney [8], using semantic orientation of phrases with adjectives/adverbs for review classification. The approach first extracts phrases with adjectives/adverbs; the phrase's semantic orientation is estimated using PMI-IR; depending on average semantic orientation of phrases the review is either recommended (Thumbs up) or not recommended (Thumbs down). Experiment used 410 reviews on various topics leading to an average accuracy of 74%.

Chaovalit and Zhou [9] investigated a two approach movie review mining; machine learning and semantic orientation. Approaches were adapted to movie review domain for comparison. Results are comparable to or even better than earlier findings. It was also found that movie review mining is more challenging than other review mining types. Movie review mining challenge is that factual information is mixed with real life review data. Also, ironic words are used for writing movie reviews. Further work to improve existing approaches was suggested.

Clasteretal[10] explored movie sentiment in Twitter microblogs. A multi-knowledge based approach using, Self-organizing Maps and movie knowledge to model opinion across multi-dimensional sentiment space was suggested. A developed visual model expresses sentiment vocabulary taxonomy which was applied to test data model. Results revealed the proposed visualization's effectiveness in Twitter sentiment mining.

Jotheeswaran et al [11] proposed feature set extraction from movie reviews. Inverse document frequency is computed and feature set reduced using Principal Component Analysis. Pre processing's effectiveness is evaluated using Naive Bayes and Linear Vector Quantization.

Valarmathi and Palanisamy [12] analyzed a method to create exclusive lists from a document's extracted words. Corpus of words created after exclude list was based on Singular Value Decomposition (SVD) scores. Classification and Regression Trees (CART) and Bayes Net with 10 fold cross validation determined classification accuracy. Obtained output was 76% and 78.667% respectively.

Kabinsinghaetal [13] investigated movie ratings. Data mining was applied to movie classification. Movies are rated into PG, PG-13 and R in the prototype. The 240 prototype movies from IMDb (http://imdb.com) were used. Data was divided into testing and training set with four fold cross validation. Among various movie attributes like actors, actress, directors, budget, genre and producers, total number of selected attributes was 8 depending on movie genres and words used in movies corresponding to a decision used by most film rating organizations. The prototype is based on the Weka used decision tree (J48) (http://www.cs.waikato.ac.nz/ml/weka/). Experiments recorded 80%-88% precision for all tested rating.deFreitas and Vieira [14] proposed evaluation methods to recognize polarity in Portuguese user produced reviews based on features described in the domain ontology's.

III. METHODOLOGY

In this paper, 400 reviews from the IMDb data set were chosen randomly with the same number of positive and negative reviews and features extracted using stemming, stop words, and inverse document frequency. Extracted features were classified with Naive Bayes, FURIA and RIDOR.

Stemming refers to root word origins. For example, Search is the root stem for words like Search, Searching, and Searches [15]. A general stop word list for words serving no retrieval purpose, but used frequently in composing documents are developed for 2 reasons: The match between query and the document is based on good indexing terms. So, documents retrieval containing words like "be", "your" and "the" in the corresponding request is not an intelligent search strategy. Non-significant words represent noise and can damage retrieval performance as they fail to discriminate between relevant and non relevant documents. Secondly, it reduces inverted file size in a 30% to 50% range [16].

Information retrieval uses Inverse document frequency (IDF) which is defined as $log2df_w/D$, where D is collection documents number and df_w is document frequency, the number of documents having w. There is a good relationship between document frequency df_w , and word frequencyf_w.

Naïve Bayes classifier

Naïve Bayes classifiers are Bayes theorem based statistical classifiers using a probabilistic approach to predict data class by matching data to class with highest posterior probability.

$$P(C_i|V) = \frac{P(V|C_i)P(C_i)}{P(V)} \tag{1}$$

Where $V = (v_1, ..., v_n)$ is the document represented in n-dimensional attribute vector and $C_1, ..., C_m$ represents m class. It is computationally expensive to compute $P(V|C_i)$. To reduce computation, Naïve assumption of class conditional independence is made [17]. Thus,

$$P(V|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \tag{2}$$

FURIA

FURIA learns rules using greedy approach implementing separate and conquer strategy. Rules are learnt for first m-1 classes, beginning with the smallest rule. The instances covered by formed rules are removed from training data after the rule is learnt with this format being adapted till no instances from target class are left. This procedure is repeated for all classes.

The rule growing process is achieved through the use of a propositional version of First Order Inductive Learner (FOIL) algorithm. Rule is initiated with an empty conjunction and adds features/ selectors till the rule covers no more negative instances. The next prospective feature is chosen so that it maximizes FOIL's information gain criterion (IG), a measure of rule improvement in comparison to default rule for the target class given by

$$IG = P_r * \left(\left(log_2 \frac{P_r}{P_r + n_r} \right) - log_2 \left(\frac{P}{P + N} \right) \right) \tag{3}$$

Where,

P_r and n_r represent positive and negative instances number covered by the rule under growing phase while P and n represent positive and negative instances number covered by default rule.

A fuzzy rule is got by replacing intervals with fuzzy intervals, namely fuzzy sets with trapezoidal membership function.

Such a fuzzy interval is specified by four parameters and written $I^F = (\varphi^{s,L}, \varphi^{c,L} \varphi^{c,U} \varphi^{s,U})$:

$$I^F(v) = \begin{cases} 1 & \varphi^{c,L} \le v \le \varphi^{c,U} \\ \dfrac{v - \varphi^{s,L}}{\varphi^{c,L} - \varphi^{s,L}} & \varphi^{s,L} < v < \varphi^{c,L} \\ \dfrac{\varphi^{s,U} - v}{\varphi^{s,U} - \varphi^{c,U}} & \varphi^{c,U} < v < \varphi^{s,U} \\ 0 & else \end{cases} \tag{4}$$

Where $\phi^{c,L}$, and $\phi^{c,U}$ are fuzzy sets membership's lower and upper bounds. A fuzzy rule is characterized by its core [ϕ^{cL},ϕ^{cU}] and support [$\phi^{s,L}$,$\phi^{s,U}$]. It is valid inside the core and invalid outside the support; the validity drops in gradually in-between [18].

RIDOR

Initially,Ripple Down Rule Learner generates default rule and exceptions are generated for it with the lowest (weighted) error rate. Then it generates "best" exceptions for every exception. Hence it carries out a tree-like exceptions expansion and its leaf gets only default rule without exceptions.

Five inner classes are defined in Ridor node class, which implements a node in Ridor tree based on default class and its exception rules.RidorRule class implements a single exception rule with Reduced Error Pruning (REP). The remaining 3 classes are only used in RidorRule namely Antd, Numeric Antd and NominalAntd.NumericAntd and Nominal Antdare the subclasses of the abstract Antd class to implement corresponding abstract functions. These subclasses implement functions related to antecedent with a nominal attribute and numeric attributes [19].

IV. RESULT AND DISCUSSION

The experiments are conducted to evaluate the accuracy, precision, recall and Root Mean Square Error (RMSE). When the number of example predicted correctly which are valid is termed True Positive (TP), the number of example predicted correctly which are invalid is termed True Negatives (TN), the number of example predicted incorrectly which are valid is termed FP (False Positive), and the number of example predicted incorrectly which are invalid is termed FN (False Negative), then accuracy is calculated as follows:

$$\text{accuracy} = \frac{(\text{TP+TN})}{(TN + FN + FP + TP)} \tag{5}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

The RMSE is given as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_{obs,i} - X_{model,i})^2}{n}} \tag{8}$$

Where X_{obs} is observed values and X_{model} is modelled values for n examples.

Classification accuracy, RMSE, Precision and recall for the three classification techniques used are shown in Table 1.

Table 1: Summary of results

Techniques Used	Classification Accuracy %	RMSE	Avg Precision	Avg Recall
RIDOR	76	0.4899	0.764	0.76
Naïve Bayes	85.25	0.3764	0.853	0.853
FURIA	92.25	0.2312	0.926	0.923

The results of classification accuracy with various methods are depicted in Figure 1.

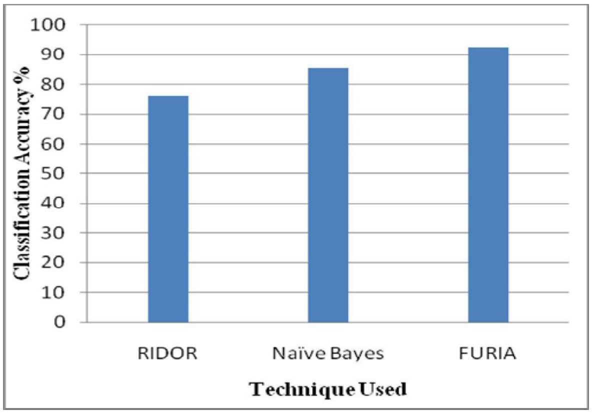

Fig 1 classification accuracy for different techniques used

The above figure shows that the classification accuracy is better in FURIA when compared with other methods.
The result of RMSE obtained with various methods is shown in Figure 2.

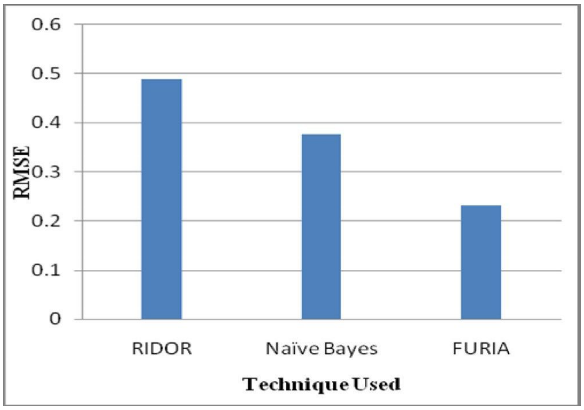

Fig 2 RMSE value for different techniques used

RMSE is comparatively lower in FURIA than other methods as mentioned in the above figure 2. Average precision and average recall obtained by different techniques are shown in Figure 3.

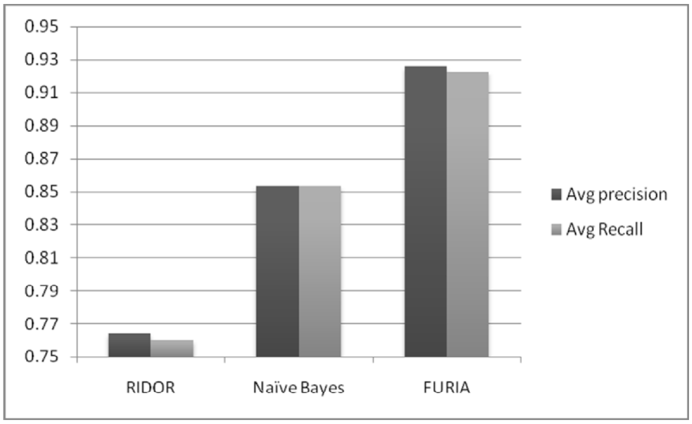

Fig 3 precision and recall for different techniques used

From Fig 3 it is observed that FURIA obtains high classification accuracy of 92.25% compared to other classifiers used.

V. CONCLUSION

To prove the efficiency of FURIA method for classification accuracy through opinion mining, we have compared three methods RIDOR,Naïve Bayes and FURIA.Further it can extent to improve the performance of the system using feature reduction technique. Also 400 reviews were randomly selected from IMDb dataset and feature extracted using stop word, stemming and IDF. The performance of FURIA classifier is better than Naïve Bayes by 8.21 % and by 21.71 compared to RIDOR.

REFERENCES

[1] Xia, Y. Q., Xu, R. F., Wong, K. F., and Zheng, F. “The unified collocation framework for opinion mining”, Proc. IEEE Int. Conf. on Machine Learning and Cybernetics, Vol. 2, pp. 844-850, 2007.

[2] RushdiSaleh, M., Martín-Valdivia, M. T., Montejo-Ráez, A., and Ureña-López, L. A. “Experiments with SVM to classify opinions in different domains”.Expert Systems with Applications, Vol.38,No.12,pp.14799-14804,2011.

[3] Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., and Gordon, J. “Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets”, Proc. Int. Conf. on on Advances in Artificial Intelligence,Vol.1,pp.1-14,2012.

[4] Chen, H., and Zimbra, D. “AI and opinion mining”, Intelligent Systems, IEEE, Vol.25,No.3, pp.74-80,2010.

[5] Internet movie database: http://www.imdb.com/

[6] Peralta, V. “Extraction and Integration of MovieLens and IMDbData”,Technical Report, LaboratoirePRiSM, Université de Versailles, France,2007.

[7] Gupta, V., and Lehal, G. S. “A survey of text mining techniques and applications”. Journal of Emerging Technologies in Web Intelligence, Vol.1,No.1, pp.60-76,2009.

[8] Turney, P., and Littman, M. L. “Unsupervised learning of semantic orientation”, Hundred-billion-word corpus,2002.

[9] Chaovalit, P., and Zhou, L. “Movie review mining: A comparison between supervised and unsupervised classification approaches”, Proc. IEEE Int. Conf. on System Sciences, pp. 112c-112c,2005.

[10] Claster, W. B., Hung, D. Q., and Shanmuganathan, S. “Unsupervised artificial neural nets for modeling movie sentiment”. Proc. IEEE Int. Conf. on Computational Intelligence, Communication Systems and Networks, pp. 349-354,2010.

[11] Jotheeswaran, J., Loganathan, R., and MadhuSudhanan, B. “Feature Reduction using Principal Component Analysis for Opinion Mining”, International Journal of Computer Science and Telecommunications ,Vol.3, No.5,pp.118-121, 2012.

[12] Valarmathi, B., and Palanisamy, V. “Opinion Mining Classification Using Key Word Summarization Based on Singular Value Decomposition”,International Journal on computer science and Engineering, Vol.3,No.1,pp.212-215,2011.

[13] S.Kabinsingha,S., Chindasorn,C., and Chantrapornchai.A. “Movie Rating Approach and Application Based on Data Mining”, International Journal of Engineering and Innovative Technology ,Vol. 2, No.1, pp.77-83,2012.

[14] de Freitas, L. A., and Vieira, R. “Ontology-based Feature Level Opinion Mining for Portuguese Reviews”, Proc. Int. Conf. on World Wide Web companion,pp.367-370,2013.

[15] Lovins,J.B.“ Development of a stemming algorithm”,Mechanical Translation and Computational Linguistics, vol.11,No.1,2,pp. 22-31,1968.

[16] Savoy, J. “A stemming procedure and stopword list for general French corpora”, Journal of the American Society for Information Science, Vol.50,No.10,pp. 944-952,1999.

[17] McCallum, A., and Nigam, K. “A comparison of event models for naive bayes text classification”, In AAAI-98 workshop on learning for text categorization,Vol.752, pp.41-48, 1998.

[18] Rahman, M. M., and Davis, D. N. “Machine Learning Based Missing Value Imputation Method for Clinical Dataset” IAENG Transactions on Engineering Technologies,Volume 229, pp 245-257, 2013.

[19] Devasena, C. L., Sumathi, T., Gomathi, V. V., and Hemalatha, M. “Effectiveness Evaluation of Rule Based Classifiers for the Classification of Iris Data Set”, Bonfring International Journal of Man Machine Interface, Vol.1(Special Issue),pp.05-09,2011.