International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 02 , February 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

WEIGHTING INDIVIDUAL OPTIMAL FEATURE SELECTION IN NAIVE BAYES FOR TEXT CLASSIFICATION

Madhuri B. Bhusare[1], Chandrakant R. Barde[2]

[1]*Department of Computer Engineering G.E.Ss R.H.Sapat C.O.E.M.S and R,Nashik,(India)*

[2]*Department of Computer Engineering G.E.Ss R.H.Sapat C.O.E.M.S and R, Nashik,(India)*

ABSTRACT

In this massive amount of data, data is too vast so that text categorization is important issue. With the help of previously organize set of documents and classes we can automatically classify data. The main aim of this paper is find out patterns which improve speed as well as accuracy of text mining in this digital era. For that we have to select better feature will improve performance classification and the illuminate the inter-pretention. For good organization here first use kullback Leibler Divergence and Jefferies Divergence as binomial representation that matter type I and type II error, So establish another measure i.e Jeffreys Multi Hypothesis Divergence for multiclass classification and found two method of feature selection first is maximum discrimination and second is maximum discrimination X^2. Further to improvement the bias capacity develop feature selection algorithms by weighting each distinct feature.

Index Terms: Feature Selection, Feature Reduction, Jeffreys Divergence, Kullback-Leibler Divergence.

I INTRODUCTION

In this edge of digital era with increasing availability of data. All Text document is in electric form, if it is great importance to label the content with predefine Set of document categories in an automated text categorization in previous machine learning algorithm have been develop to address this challenging task by formulating it as classification problem. Generally an automatic text classifier is strengthen with learn-ing process from a set previously labelled document, So the documents and data are needed to be represented in a proper way. That it is well suitable for general machine learning process the most widely used representation some amount of words. Challenging task in text categorization is finding patterns from multi-dimensional data. On one hand this huge amount of Increasing data terms will introduce computational complexity of learning process. It contain lots of low frequent features. When feature selection is became important task in this digital era huge amount of data is available in that irrelevant data or noisy data.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

That time features selection is necessary. May hurts prognosticative performance of classification in text categorization. A common approach of feature reduction is to introduce those features that are multi-type. One feature can act as multiple feature. Too many no of dataset are available, so this feature can compare with this and performbetter text categorization with small amount features. The purpose of selecting this paper is that in previous work only a subset of original feature are selected as input to the learning process. In this paper we present a feature selection method. Which ranks the original feature, aiming to maximize the discriminative performance for the text categorization. When the Naive Bayes classifiers are used as a learning algorithm.

1.1. Naive Bayes

It mostly used for text categorisation because it is more efficient and easily understandable [10], [11], [15]. For reducing features we add some pre-processing steps in naive Bayes .It provide competitive performance compared to other data driven classification methods. Sometimes result of naive Bayes get poor because of some unrealistic assumptions. It applies Bayes theorem with the naive assumption it means any pair of features are independent for class. Maximum posteriori take decision in naive Bayes. It is used in many data mining application. It contain three model first Bernoulli model second is Multinomial it shows real-life benchmark[21], [22]and third is land position model and its framework result is Bernoulli naive Bayes, Multinomial naive Bayes and Position naive Bayes.

Is old method for selecting feature. The feature selection is like logistic regression. It can done two measure task classification and clustering. Classification is done on previously define classes on that basis we can introduce new classes. Some algorithms can use in machine learning KNN [3] k is user define parameter it is the feature space. Feature can extract from dataset. Classification is done according to KNN[5] and SVM [7]. For new class instance is define according to the place of feature space related to hyper surfaces. SNM mention two properties that space between different classes the algorithm place on hyper surface in cantered as possible as and mount data instances into multi-dimensional space, so it will classify into different classes from each other. Clustering is nothing but finding groups of data instances so that data in one cluster belowthat cluster characteristics. K-mean [8]. Cluster representative name as centroid it show the characteristics of that cluster.

1.2.Knowledge Extraction

It is process of identifying valid, potential usfull which is easily understandable that extract from dataset that data is unstructured form and must be readable. The degree of extracting valid data is nothing but automation. Data mining is used for identifying pattern from large no of dataset that contain artificial intelligence and machine learning. It is analysis step for knowledge discovering. Data mining is extract data from group of cluster, unusual record, association rule of mining. Extracting pattern by Bayes method is early. Data stored properly in cluster and assign indexing in database is nothing but database management.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

II REVIEW OF LITERATURE

This paper introduces support vector machine for text categorization. It provides both theoretical and empirical evidence that SVMS are very well suitable for text categorization the theoretical analysis concludes that SVMS acknowledge the particular properties of text. i.e. a. High dimensional features spaces. b. Few irrelevant features c. Sparse instance vector [1]. Our system develops a new approach towards the automatic text categorization. Model learn and then it is used for classifying the feature document. This categorization approach is derived from a machine learning paradigms which known as dynamic learning. In which we retrieve advance document is known as retrieval feedback [2].

In most of resent text categorization research focuses on addressing specific issue in text categorization [Ex. Feature Selection, Dimensional Reduction] Very few new approaches are be devise [4].

In recent days the problem of ranking has gained much attention in machine learning ranking method s may filter feature to reduce dimensionality of the feature space this is very much effective for classification method that do not have any inherent feature selection built in Ex. Nearest neighbour method, some type of neural network[5].

The system have demon striated that hierarchical neural lan-guage model can actually out perform its non-hierarchical counter parts and achieve state of art performance the main motto is to making a hierarchical model perform well is using to carefully constructed hierarchy over words[6].

This paper describe paragraph vector, an unsupervised learning algorithms That learn vector represent for variable length pieces of text such as sentences and documents the vector representation are learned to predict the surrounding word in context. Sampled from the paragraph [12].

For successful information retrieval, the Naive Bayes model has been used, producing some of the best result. In the recent comparisons of learning method for text categorisation have been somewhat less favourable nave bayes model ,while still showing them achieve respectable effectiveness[13].

The recent increasing of system that hierarchically organize massive amounts of text based documents calls algorithms that hierarchically categorization new document as they come in. The paper describes an approach which utilizes the ex-isting reach hierarchical structure in order to facilitate this process [14].

This paper present an extensive comparative study for feature selection matrix for high dimensional domain of text classification. While focusing on support vector machine and to class problems typically with high class skew, it has revelled the surprising performance of a new feature selection matrix by normal separation [16].

By achieving the equivalent Feature Selection within the con-text of Well-founded bastion, the losso approach provides the great scope for Domain knowledge. Combining with learning from training data. Simplest care would be to give rare world prioress with higher variance, reflecting that they have higher contain than more common words [18].

The paper has compared the theory and practice of to different first order probabilistic classifier, both of which make the Naive Bayes assumption. The multinomial model is found to be almost uniformly better than multivariate born Olly model in impaired result. On five real word corpora. We find that the multi nominal model reduces error by an

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

average of 27% and sometimes by more than 50% [20].

In this paper we have discuss and evaluated experimentally in span filtering Context 5 different versions of nave Bayes classifier this investigation included two version of nave Bayes that have not been used widely in spam filtering literature namely flexible Bayes and multinomial nave Bayes with bullion attribute[22].

Our system introduce a new filtering measure for feature selection in text categorization. Which have simple expression in terms of appearance of the word im the different documents. It also shows that these measures have interesting property [24]. The system proposes a distributed Naive Bayes text classification model with wait enhancing method. These module assumes a documents is generated by multivariate.distribution module so the system suggest per document term frequency normalization to estimate the distribution parameters [25].

Text categorization has found many application in data classification, spam detection, data analysis also have various algorithms. Nave Bayes is most popular and simple algorithm. Classification algorithm usually require training data. Which causes space complexity. So improve capacity of classifier algorithm to give good performance on relatively low training data.

III PROPOSED SYSTEM

3.1 Problem Statement

- Feature selection method can ranks and also assign weight to the original features, so maximize the discriminative performance for text categorization, when naive Bayes classifiers are used as learning algorithms.
- Efficient approach to rank the order of features to approximately produce the maximum JMH divergence. The theoretical analysis shows that the JMH divergence is monotonically increasing when more features are selected.

3.1.1 Feature Reduction:

We can reduce data by eliminating feature to obtain accuracy in data mining and remove noise. It also compress the data for efficient retrieval and storage purpose. Data are pre-process for good pattern mining and machine learning.

- Diversion Measures for Binary Hypothesis Testing: In probability theory and information theory the Kullback-Leibler divergence also called discrimination information.
- Jeffreys Multi Hypothesis Divergence: [29][32]first generation to next multi hypothesis divergence uses scheme that one verse all. [33],[34]It can also calculate N binary hypothesis testing detectors. Jensen Shannon used in calculating the similarity between two properties. Divergence is the one that can be used to measure multi-distribution divergence, In that divergences of each individual distribution with a reference distribution are computed and sum together.

3.1.2 Rank Feature Index:

Search engine can calculate the relevance rank. Ranking model has collection of ranking features to calculate the rank score that document. It can take data from the search index from that document.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

- Weighting to feature In that each feature in a document is assign a feature weight based on a weighting scheme.

3.1.3 Text Categorization Naive Bayes Classification:

Naive Bayes is a simple and efficient technique for classification. It assign class labels to data instances, represented as feature values, where the class labels are drawn from some finite set of document. All naive Bayes classifiers initialise that the value of a any single feature is not dependent of the value of each other. In given set of document Ex, a fruit may be considered to be an orange its colour is orange, round, and about 8 cm in diameter. For classification considers every features to calculate independently to prove that this fruit is an orange, otherwise its not orange. Correlations between the colour, roundness and diameter are nothing but the features of Orange fruit.[13], [14] Nave Bayes is use in many data mining application. Naive Bayes classifiers can be better for supervised learning. In many practical applications, parameter estimation for naive Bayes models uses the method of maxi-mum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods. It worked effectively in many complex real-world situations. Analysis of the Bayesian classification is high sometimes it get lowinactive performance for text categorization, when naive Bayes classifiers are used as learning algorithms

IV SYSTEM ARCHITECTURE

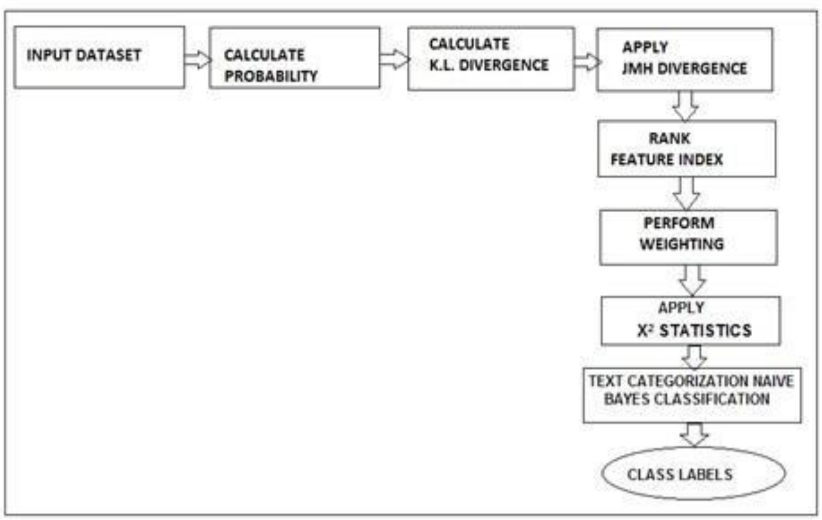

Fig. 1 System Architecture

1.Input data set

Reuters dataset contain 21578 documents, in that 90 types and 200 tokens.9603 training article and 3295 test article and 118 categories. Average document has 124 classes in that at least one category.

4.1 Pre-processing:

4.1.1 Feature Selection:

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

It can avoid Over fitting to achieve good generalize ability. It omit low frequent feature [17], [26],[27].for feature selection popular method is filtering in that assigned rank every feature and those have highest rank that only consider others are removed. It is vary simple and have low computational cost.

1. Document Frequency: No of Document in which term is occurs is simple and effective feature selection approach.[28]TF-IDF is also consider to compute importance of features.

2. Mutual information : It calculates the mutual de-pendency of two variables. It also use for common feature extracting method.

3. Information Gain : [17], [31]It find out weather that term present or not in that document.

Chi Square Statistics : It compute lack of independence between the term tk and Ci which configure by Chi square distribution.[33] KL divergencebecause of some irrelevant feature consideration. Still it good in comparison with other classification algorithms. It is good because it require small amount of training set of data to classify or estimate properly.

4.1.2 Probability Calculation

In Naive Bayes classification is based on Bayes theorem with not dependence consideration among two predictors. The Naive Bayes classification often does well and is widely used because it often outperforms more sophisticated classification methods. Bayes theorem provides a way of computing the posterior probability.

V SYSTEM ANALYSIS

To classify low level features according to their appearances in the higher level text mining and their specificity. For that it introduces weighting feature method can select irrelevant document. Relevance features for any topic, normally specific terms are very important in order to distinguish the topic from other topics. It show that using only specific terms is not good enough to improve the performance of relevance feature discovery because user information needs cannot simply be covered by documents that contain only the specific terms. So, the best way is to use the specific terms mixed with some of the general terms. Issue in the evaluation time. For improve the effectiveness, the Relevance Feature Discovery used to find irrelevant documents in the training set in order to remove the that irrelevant documents. It gives expected performance but it require manual testing. Issue of using irrelevant documents is how to collect a suitable set of irrelevant documents hence a very large set of negative samples is obtained. For example, When user Search any term on Google search engine it can return thousands of documents; From that, only a few of those documents may be of interest to a Web user. Obviously, it is not efficient to use all of the irrelevant documents.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Algorithm Steps for Feature Selection :

Input: Term Document Matrix (TM)(D*W) D indicates No. of documents
W indicates no. of words an entry TMij indicates the corresponding tf-idf.
No. of Clusters (nc), Default Value 0
Output: Feature Set F initially empty set

Step 1: Apply feature selection algorithm based on chi-squared(CH) the entire term document calculate chi-squared (CH)value according to every word.
Step 2: Select words whose CH value more than default value.
Step 3: New term document matrix (TM) consists of only those important words as selected in Step 2.
Step 4:New term document matrix (M) and each row represents a word. The transposed matrix is denoted by N.
Step 5: Create nc clusters on N.
Step 6: Most representative words are retrive from every cluster and those are closest to the clustering centre that points add one by one to F such n(F)=nc.
Step 7: Euclidian norm calculate every point in a cluster, between the point and the center. The one nearest to the center is selected

Algorithm steps to perform Weighting On Feature:
Input : Set of terms T

Dr : Set of relevant Document Di : Set of irrelevant Document
Sp : Set of discovered closed sequential patterns for all documents.
DPr : Set of discovered pattern of Dr Dpi : Set of discovered pattern of Di.
Ts : Specific feature ; Tg : General Feature P : Average size of pattern.

Step 1:Weighting Feature function assign weight to every term
Step 2:Feature Clustering is apply for classification.
Step 3:Calculate Sup function and Spe function on every term.
Step 4:Calculate $D_S$upfunction
Step 5:Perform feature clustering to classify terms into two categories
-Specific feature -General feature
Step 6:Calculate the weight for every term using weighting function.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Algorithm Steps for nave Bayes Classification

Step 1: First create the frequency table for giveb data set.

Step 2: Create Likelihood table by finding the probabilities.

Step 3: Naive Bayesian equation is calculate the posterior probability for each class. Then those classes have highest posterior probability is the outcome of prediction.

5.1 Mathematical Model

S : { fD, KL, JHM, RF, AX, NB, CI }

Where,

S = is a System, D = Input dataset,

KL = Calculate K L Divergence, JHM = JHM Divergence,

RF = Rank Feature Index, AX = Apply X^2,

NB = Nave Bayes

CL =Classification

Y : { KL, JHM, RF, AX, NB, CI }

D: d1, d2,., dn Set of Web Data.

F: f1, f2,.., fn Functions for Feature Selection. Y is a set of techniques use in our Application.

State(S)

At client side : text classification from collection of document.

At server side : Perform Calculation.

End State(E)

At client side : class label get and do verification.

At Server side : According to user requirement send classification results.

Input (I) : Data set send to the system which are further get classified .

Output (O) : According to user requirement data set get classified on system.

Function (F): Probability calculation, Computing KL Divergence and JMH Divergence, Rank Feature, Weighting Feature

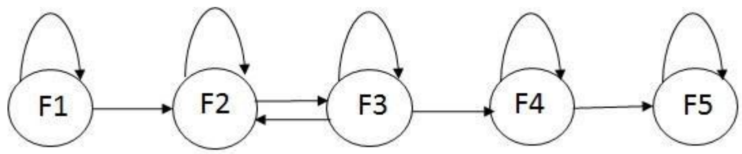

Fig.2 State Transition Diagram

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Where,

F1 : Calculate K-L Divergence

F2 : JHM Divergence

F3 : Rank Feature Index

F4 : Apply X^2 Statistics

Naive Bayes Classification

5.1 Limitations

1. Classification algorithm require training data. Which causes space complexity.
2. If increase the performance of classifier, Then training time also increases of data set.

VI EXPERIMENTAL SETUP AND RESULTS

Table1 Shows the comparison of existing systems with proposed system. there are various existing system with there classifier is compare with proposed system .Proposed System is distinguished with the high value of F-measure ,accuracy, precision, recall from all existing system. Proposed System also represent the higher average of weighting feature selection algorithm with better accuracy and precision,F-measure etc. hence table shows the proposed system with the high average values of all classifier.

Table 1.Comparison of F-measure ,accuracy, precision, recall using existing with proposed system.

Classificatio	F-measure	Accuracy	Presision	Recall
KNN	82.3	78.5	86.65	85.88
Kmeans	75.48	84.45	78.56	78.66
Multinomial	85.56	79.6	74.89	84.54
Bernoli NB	88.5	82.68	84.78	83.43
SVM	78.9	76.68	88.65	87.36
Wfeature	90.6	94.43	95.8	92.67

Table1 Shows the comparison of existing systems with proposed system. there are various existing system with there classifier is compare with proposed system .Proposed System is distinguished with the high value of F-measure ,accuracy, precision, recall from all existing system. Proposed System also represent the higher average of weighting feature selection algorithm with better accuracy and precision,F-measure etc. hence table shows the proposed system with the high average values of all classifier.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

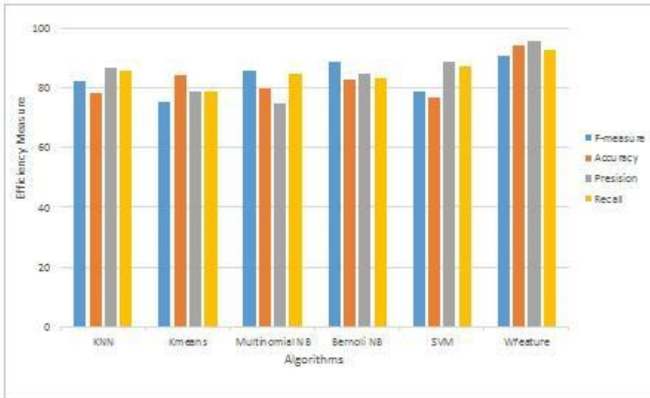

Fig. 4. Graph representation of existing and Propose system.

The proposed system has introduce new feature selection on the basis of information measures for Naive Bayes classifiers, aiming to select the features that offer the maximum discriminative capacity for text classification, Thus we can achieve the optimal feature selection in Naive Bayes for text categorization.

We had carry out experiments on the dataset mentioned in paper (Reuters) when naive Bayes and SVM are used as classifiers. To compare the performance of these feature selection methods, we evaluate the classification-measure accuracy and precision, recall metric of these classifiers with different number of features ranging from 10 to 2000.

VII CONCLUSION

Feature selection approaches based on the information measures for naive Bayes classifiers, aiming to select the features that offer the maximum discriminative capacity for text classification. It also develop feature selection algorithms by weighting each individual features, aiming to maximize the discriminative capacity

VIII ACKNOWLEDGMENT

I would like to express gratefulness to P.G Dept. of Com-puter Engineering,GESs R.H. Sapat C.O.E.M.S and R, Nashik.

REFERENCES

[1] T. Joachims, Text categorization with support vector machines: Learning with many relevant features, in Proc. Eur. Conf. Mach. Learn., 1998, pp. 137142.

[2] W. Lam, M. Ruiz, and P. Srinivasan, Automatic text categoriza-tion and its application to text retrieval, IEEE Trans. Knowl. Data Eng., vol. 11, no. 6, pp. 865879, Nov./Dec. 1999.

[3] I. Witten and E. Frank, Data Mining: Practical Machine LearningTools and Techniques, Second Edition, ser.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

The Morgan KaufmannSeries in Data Management Systems. Elsevier Science, 2005.

[4] H. Al-Mubaid and S. Umair, A new text categorization technique using distributional clustering and learning logic, IEEE Trans. Knowl. Data Eng., vol. 18, no. 9, pp. 11561165, Sep. 2006.

[5] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed. Prentice Hall, 1998.

[6] H. Liu and L. Yu, Toward integrating feature selection algo-rithms for classification and clustering, IEEE Trans. Knowl. Data Eng., vol. 17, no. 4, pp. 491502, Apr. 2005.

[7] V. N. Vapnik, The Nature of Statistical Learning Theory. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

[8] J. B. MacQueen, Some methods for classification and analysis of mul-tivariate observations, in Proceedings of 5th Berkeley Symposiumon Mathematical Statistics and Probability, 1967, pp. 281-297.

[9] A. Mnih and G. E. Hinton, A scalable hierarchical distributed language model, in Proc. Adv. Neural Inform. Process. Syst., 2009, pp. 1081-1088.

[10] D. D. Lewis, Naive (Bayes) at forty: The independence assump-tion in Inform. retrieval, in Proc. Eur. Conf. Mach. Learn., 1998, pp. 415

[11] D. Koller and M. Sahami, Hierarchically classifying documents using very few words, in Proc. Int. Conf. Mach. Learn., 1997, pp. 170-178.

[12] Q. V. Le and T. Mikolov, Distributed representations of sentences and documents, arXiv preprint arXiv:1405.4053, 2014.

[13] D. D. Lewis, Naive (Bayes) at forty: The independence assumption in Inform. retrieval, in Proc. Eur. Conf. Mach. Learn., 1998, 415.

[14] D. Koller and M. Sahami, Hierarchically classifying documents using very few words, in Proc. Int. Conf. Mach. Learn., 1997, 170-178.

[15] Y. H. Li and A. K. Jain, Classification of text documents, Com-put. J., vol. 41, no. 8, pp. 537-546, 1998.

[16] G. Forman, An extensive empirical study of feature selection metrics for text classification, J. Mach. Learn. Res., vol. 3, 1289-1305, 2003.

[17] Y. Yang and J. O. Pedersen, A comparative study on feature selection in text categorization, in Proc. Int. Conf. Mach. Learn., vol. 97, 1997, pp. 412-420.

[18] A. Genkin, D. D. Lewis, and D. Madigan, Large-scale Bayesian logistic regression for text categorization, Technometrics, vol. 49, no. 3, pp. 291-304, 2007. B. Tang and H. He, ENN: Extended nearest neighbor method for pattern recognition [research frontier], IEEE Comput. Intell. Mag., vol. 10, no. 3, pp. 5260, 2015.

[19] A. McCallum and K. Nigam, A comparison of event models for naive Bayes text classification, in Proc. AAAI Workshop Learning Text Categorization, 1998, vol. 752, pp. 4148.

[20] S. Eyheramendy, D. D. Lewis, and D. Madigan, On the naive Bayes model for text categorization, in Proc. 9th Int. Workshop Artif. Intell. Statist., 2003, pp. 332-339.

[21] V. Metsis, I. Androutsopoulos, and G. Paliouras, Spam filtering with naive Bayes-which naive Bayes? in Proc.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Conf. Email Anti-Spam, 2006, pp. 2728.

[22] E. F. Combarro, E. Montanes, I. Diaz, J. Ranilla, and R. Mones, Introducing a family of linear measures for feature selection intext categorization, IEEE Trans. Knowl. Data Eng., vol. 17, no. 9, pp

[23] S. Kay, Q. Ding, B. Tang, and H. He, Probability density function estimation using the EEF with application to subset/feature selection, IEEE Trans. Signal Process., vol. 64, no. 3, pp. 641651, Feb. 2016.

[24] B. Tang, H. He, P. M. Baggenstoss, and S. Kay, A Bayesian classi-fication approach using class-specific features for text catego-rization, IEEE Trans. Knowl. Data Eng., vol. 28, no. 6, pp. 1602 1606, Jun. 2016.

[25] D. Mladenic and M. Grobelnik, Feature selection for unbalanced class distribution and naive Bayes, in Proc. Int. Conf. Mach. Learn., 1999, vol. 99, pp. 258-267.

[26] G. Salton, A. Wong, and C.-S. Yang, A vector space model for automatic indexing, Commun ACM, vol. 18, no. 11, pp. 613620, 1975.

[27] J. Lin, Divergence measures based on the Shannon entropy, IEEE Trans. Inf. Theory, vol. 37, no. 1, pp. 145151, Jan. 1991.

[28] D. Mladenic and M. Grobelnik, Feature selection for unbalanced class distribution and naive Bayes, in Proc. Int. Conf. Mach. Learn., 1999, vol. 99, pp. 258267.

[29] M. F. Caropreso, S. Matwin, and F. Sebastiani, A learner-independent evaluation of the usefulness of statistical phrases for automated text categorization, Text Databases and Document Management: Theory and Practice. Hershey, PA, USA: IGI Publish-ing, pp. 78102, 2001.

[30] K. Sawyer, A multiple divergence criterion for testing between separate hypotheses, Statist. Probability Lett., vol. 1, no. 1, pp. 26 30, 1982.

[31] Bo Tang, Student Member, IEEE, Steven Kay, Fellow, IEEE, and Haibo He, Toward Optimal Feature Selection in Naive Bayes for Text Categorization,IEEETrans.Knowl. Data Eng., vol. 28, no. 9, Sep. 2006.

[32] R. Rifkin and A. Klautau, In defense of one-vs-all classification, J. Mach. Learn. Res., vol. 5, pp. 101141, 2004. B. Tang and H. He, ENN: Extended nearest neighbor method for pattern recognition [research frontier], IEEE Comput. Intell. Mag., vol. 10, no. 3, pp. 5260, 2015.

[33] A. McCallum and K. Nigam, A comparison of event models for naive Bayes text classification, in Proc. AAAI Workshop Learning Text Categorization, 1998, vol. 752, pp. 4148.

[34] S. Eyheramendy, D. D. Lewis, and D. Madigan, On the naive Bayes model for text categorization, in Proc. 9th Int. Workshop Artif. Intell. Statist., 2003, pp. 332-339.

[35] V. Metsis, I. Androutsopoulos, and G. Paliouras, Spam filtering with naive Bayes-which naive Bayes? in Proc. Conf. Email Anti-Spam, 2006, pp. 2728. [36] E. F. Combarro, E. Montanes, I. Diaz, J. Ranilla, and R. Mones, Introducing a family of linear measures for feature selection intext categorization, IEEE Trans. Knowl. Data Eng., vol. 17, no. 9, pp

[37] S. Kay, Q. Ding, B. Tang, and H. He, Probability density function estimation using the EEF with application to subset/feature selection, IEEE Trans. Signal Process., vol. 64, no. 3, pp. 641651, Feb. 2016.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

[38] B. Tang, H. He, P. M. Baggenstoss, and S. Kay, A Bayesian classi-fication approach using class-specific features for text catego-rization, IEEE Trans. Knowl. Data Eng., vol. 28, no. 6, pp. 1602 1606, Jun. 2016.

[39] D. Mladenic and M. Grobelnik, Feature selection for unbalanced class distribution and naive Bayes, in Proc. Int. Conf. Mach. Learn., 1999, vol. 99, pp. 258-267.

[40] G. Salton, A. Wong, and C.-S. Yang, A vector space model for automatic indexing, Commun ACM, vol. 18, no. 11, pp. 613620, 1975.

[41] J. Lin, Divergence measures based on the Shannon entropy, IEEE Trans. Inf. Theory, vol. 37, no. 1, pp. 145151, Jan. 1991.

[42] D. Mladenic and M. Grobelnik, Feature selection for unbalanced class distribution and naive Bayes, in Proc. Int. Conf. Mach. Learn., 1999, vol. 99, pp. 258267.

[43] M. F. Caropreso, S. Matwin, and F. Sebastiani, A learner-independent evaluation of the usefulness of statistical phrases for automated text categorization, Text Databases and Document Management: Theory and Practice. Hershey, PA, USA: IGI Publish-ing, pp. 78102, 2001.

[44] K. Sawyer, A multiple divergence criterion for testing between separate hypotheses, Statist. Probability Lett., vol. 1, no. 1, pp. 26 30, 1982.

[45] Bo Tang, Student Member, IEEE, Steven Kay, Fellow, IEEE, and Haibo He, Toward Optimal Feature Selection in Naive Bayes for Text Categorization,IEEETrans.Knowl. Data Eng., vol. 28, no. 9, Sep. 2006.

[46] R. Rifkin and A. Klautau, In defense of one-vs-all classification, J. Mach. Learn. Res., vol. 5, pp. 101141, 2004.

AUTHOR PROFILE

Madhuri B. Bhusare is pursuing the Masters in Computer from G. E. S. R. H. Sapat College of Engg., Nashik under Pune University.She has pursued her Bachelors Degree in Computer Technology from Matoshree college of engineering, Nashik under Pune University.

Prof.Chandrakant R. Barde is currently working as an As-sistant Professor in the Computer Engineering Department of GESs R. H. Sapat College of Engineering Management Stud-ies and Research ,Nashik (INDIA).He received his Masters Degree in Computer Engineering from Dr.B.A.T.U., Lonere, Raigad in 2012 and Bachelors Degree in Computer Science and Engineering from Amruthvahini College of Engineering in 2005. He has an academic experience of 11 years (since 2005).