International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 4 577 - 580

Minimum Redundancy Maximum Relevance(mRMR) Based Feature Selection Technique for Pattern Classification System

Prerna Kapoor
M. Tech. (CSE), Dept. of CSE, SRM University
NCR Campus Modinagar Ghaziabad
Uttar Pradesh, India
prernakapoor37@gmail.com

Mr Madan Singh
Assistant Professor, Dept. of CSE, SRM University
NCR Campus Modinagar Ghaziabad
Uttar Pradesh, India
madan.shishodia@gmail.com

***Abstract*—**Feature Selection is an important hurdle in classification systems. We study how to select good features by making the covariance matrix of each sample data set and extracting the features from it .Then, we try to find out the length of each sample by finding the error rate .We perform experimental comparison of our algorithm and other methods using two data sets(binary and functional) and three different classifiers(support vector machine, linear discriminant analysis and naïve Bayes).The results show that the MRMR features are less correlated with each other as compared to other methods and hence improves the classification accuracy.

Keywords- *Covariance matrix, mRMR, Feature Selection,Error Rate*

*****

I. INTRODUCTION

Features or attributes represent the characteristics of malware or benign samples. Generally a subset of features represent these characteristics. However, dataset used for analysis contain numerous attributes which increases the difficulty of any classification system.

Learning is very hard in high dimensional datasets .High computational cost is one of the important drawback in these systems. But any real-word data can be distributed uniformly in a high dimensional space. To find out the best subspace a search algorithm is needed. So this condition of finding out the best subspace with minimal classification error is called "optimal characterization". Minimal error usually requires the maximal statistical dependency of the target class on the data distribution in the subspace. This is called maximal dependency(Max- Dependency).

Maximal relevance(Max- Relevance) feature selection is the most popular approach to realize Max-Dependency. It means selecting the features that has the highest relevance to the target class. Relevance means the mutual information which is the widely used measure to define the dependency of features. So here we discuss the feature selection method based on mutual information. If suppose two features or variables say a and b are given then the mutual information is defines in term of their probabilistic density functions. In this approach the selected features are required to have the largest mutual information with the target class i.e. the largest dependency on the target class. In sequential search the top most features which has the highest values of mutual information are selected as the optimal feature subset. But it has been observed that the combination of good features does not relate to good performance in classification. So we need to reduce the redundancy among features and select the minimal redundant features. Also a pre-filtering method can be applied to group the features and thus redundant features within each sample can be removed. In this paper we proposed a framework called minimal-redundancy maximal-relevance to minimize redundancy and applied a number of steps of redundancy and relevance to select the best features for both the binary and functional data sets.

Our work in this paper focuses on the theoretical analysis of mRMR which is a way to implement maximum dependency for feature selection. We extracted the features from the other classifiers namely naive bayes, linear discriminant analysis and support vector machine and then applied this approach to reduce the error rate from the features of a particular sample. This property of mRMR to combine with the other classifiers enable us to find a combat subset of superior features at very low cost. To select mutually exclusive features we combine minimal redundancy condition with the maximal selected features.

This paper presents an approach of feature selection based on analysis of covariance matrix of sample datasets, a mutual information-based feature selection method is put forward. It is shown that for a given set of features, an optimal feature subset that has the highest sum of the correlation coefficients has the tendency to be reduced, if it meets the requirement of

the objective function, a favorable sets is finally retained, when it is omitted, the good classification efficiency is obtained. The algorithm performs elimination which minimizes the value of objective measure, a terminating criterion is given. Experiments show that the proposed algorithm performs well in eliminating redundant features while reducing the error rate for individual samples.

Rest of the paper is organized into: section II Related Work, section III System Architecture Overview, section IV Results section V Conclusion.

II. RELATED WORK

R.K.Bania (2014) proposed an extensive survey on supervised *FS* technique describing the different searching approach, methods and application areas with an outline of a comparative study. The storage capabilities and advanced in data collection has led to an information load and the size of databases increases in dimensions, not only in rows but also in columns. Data reduction (*DR*) plays a vital role as a data prepossessing techniques in the area of knowledge discovery from the huge collection of data. Feature selection (*FS*) is one of the well known data reduction techniques, which deals with the reduction of attributes from the original data without affecting the main information content. Based on the training data used for different applications of knowledge discovery, *FS* technique falls into supervised, unsupervised.

Ron Kohavi (1997) presented in his paper the relation between optimal feature subset selection and relevance. To achieve the best possible performance with a particular algorithm on the sample datasets, the selection technique should consider the interaction between the algorithm and the sample dataset. While selecting the features, an algorithm is faced with the problem of selecting a relevant subset of features while neglecting the rest. The wrapper approach searches for an optimal feature subset related to a particular algorithm and domain. The author studied the advantages and disadvantages of this approach and gave a series of improved designs. The comparison of this approach to induction without feature subset selection and to Relief, a filter approach to feature subset selection. Significant improvement in accuracy is achieved for some datasets for the two families of induction algorithms used: Decision trees and Naive-Bayes.

Eric, Michael and Richard(2001) presented a report on the successful application of feature selection methods to a classification problem in molecular biology involving only 72 data points in a 7130 dimensional space. The hybrid approach of filter and wrapper approaches to feature selection. The use of a sequence of simple filters, culminating in Koller and Sahami's (1996) Markov Blanket filter, to decide on particular feature subsets for each subset cardinality. Then comparing between the resulting subset cardinalities using cross validation they also investigated regularization methods as an alternative to feature selection, showing that feature selection methods are preferable in this problem.

III. SYSTEM ARCHITECTURE

System architecture of our feature selection work is shown in figure 1. This paper presents a novel approach of feature selection based on analysis of covariance matrix of training patterns, a correlation-based feature selection method is put forward. An objective measure is proposed and defined. It is shown that for a given set of features, a subset of features that has the highest sum of the correlation coefficients has the tendency to be reduced, if it meets the requirement of the objective function, a favorable sets is finally retained, when it is omitted, the good classification efficiency is obtained. The algorithm performs elimination, the elimination of which minimizes the value of objective measure, a terminating criterion is given. Experiments show that the proposed algorithm performs well in eliminating irrelevant features while constraining the increase in recognition error rates for unknown data.

.**The Naïve-Bayesian Classifier** uses Bayes' rule to compute the probability of each class given the instance, assuming the features are conditionally independent. It computes the likelihood that the program is malicious given the features that are contained in the sample. The main assumption in this approach is that the binary data sets contain similar features as signature etc. The sample dataset consists of a set of features F from which we want to compute a class of dataset. We define C to be a random variable over the set of classes. That is, we want to compute P (C|F), the probability that a sample is in a certain class given the samples contain the set of features F. We then apply Bayes rule and express the probability as:

$$P(C|F) = \frac{P(F|C) * P(C)}{P(F)}$$

State Vector Machine has three phases: (a) input or transformation (b) learning (c) function estimation. The input data is converted into vectors in a new mathematical space which is also called feature space. If the data is not linearly separable, then SVM uses a non-linear kernel. Support vector is determined by learning phase. The function optimisation in SVM is carried out using SMO (Written and Frank, 1999) that makes the training phase efficient. And it is because training an SVM involves solving large quadratic programming problems. Larger problems are broken into smaller ones by SMO and then solved analytically, and hence the time consuming phase of optimisation is reduced.

Linear Discriminant Analysis is a statistical technique to classify sample dataset if they are linearly separable. Thus in this method the dependent variable (class we are looking for)

and the independent variables are the features that describe the dataset. We tend to minimize the total error of classification i.e. the proportion of sample that it misclassifies as small as possible.

Minimum Redundancy Maximum Relevance is one of the most popular technique to realize maximum dependency i.e. maximal relevance(Max-Relevance)feature selection: selecting the features with the highest relevance to the target class. In feature selection process combining the best features from individual samples does not necessarily lead to good classification performance. So, by direct or indirect means we tend to minimize the redundancy and select the features with minimal redundancy(Min-Redundancy).

We tested our feature selection technique on binary and functional data sets. For these datasets we calculated mutual information by many ways and tested the performance of the selected features based on the three classifiers introduced above. In this way through this paper we presented a comprehensive study on the performance of our approach in different conditions .First of all we enter the sample datasets and apply the three classification techniques on them. After that we applied minimum redundancy maximum relevance classification technique as described above to reduce the error rate. Then we created the covariance matrix of individual samples. From these matrixes we extracted features and determined the error rate for each of the samples. The next step is to find the length of each of them and then comparing with the error rates of the samples found by other classification techniques. The results confirm that mRMR leads to promising improvement on feature selection and hence improves the classification accuracy.

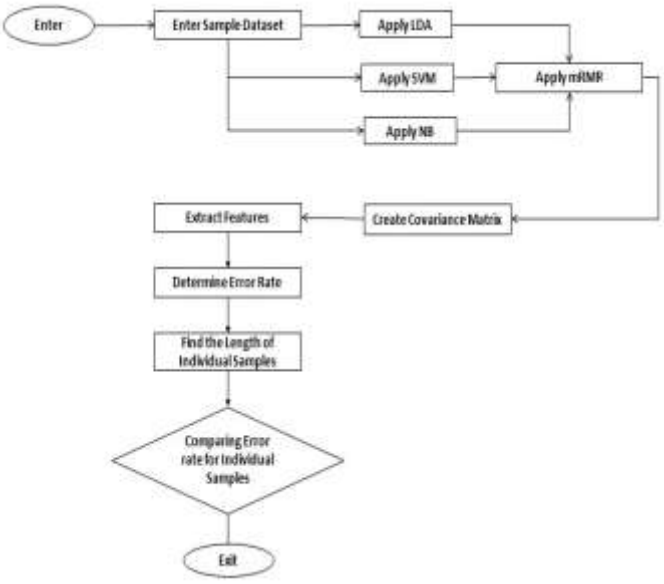

Figure1 System Architecture

IV. RESULTS

This section is divided into three parts. This section is divided into three parts. After the brief description of the sytem architecture we now present our results in the three graphs shown above.It clearly shows the comparison of three techniques for feature selection in pattern recognitin applications.

Figure1 clearly depicts the reduced error rate for individual features when mRMR is applied to naive bayes classifier.For eg ,at the 25th feature error rate reduces from 1 to 0.5.

Figure2 presents the reduced error rate for individual features when mRMR is applied to linear discriminant analysis classifier.For eg,at the 6th feature error rate reduces from 1.4 to 0.6.

Figure3 shows the reduced error rate for individual features when mRMR is applied to state vector machine classifier.For eg,at the 12th feature error rate reduces from0.8 to 0.

Hence we can conclude that the classification accuracy of mRMR method is much more better than the other three clasifiers.

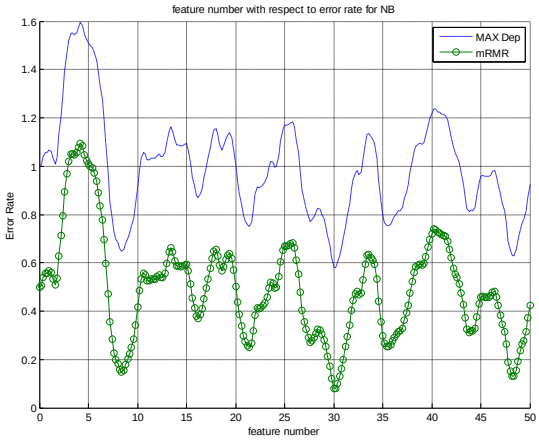

Figure2 NB and mRMR

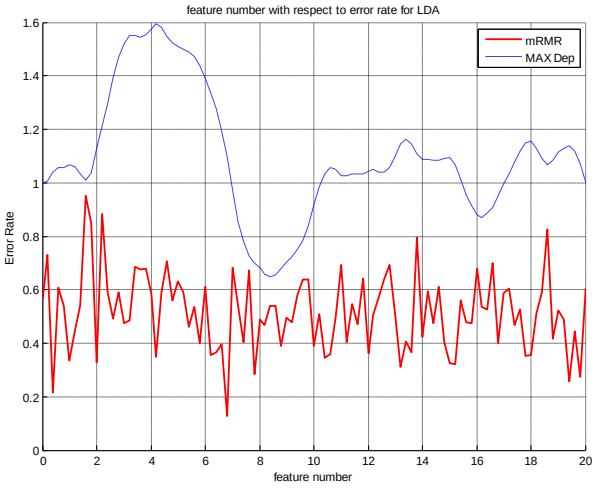

Figure3 LDA and mRMR

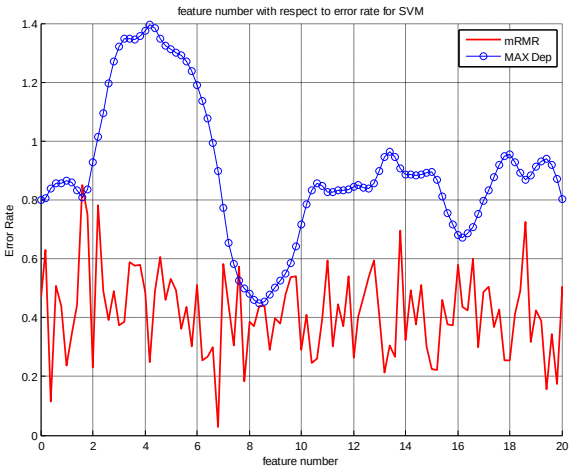

Figure4 SVM and mRMR

V. CONCLUSION

This paper has presented that minimum redundancy maximum relevance feature selection method is much more efficient than naïve bayes and linear discriminant analysis method.

This paper has presented that minimum redundancy maximum relevance feature selection method is much more efficient than naïve bayes and linear discriminant analysis method.

This paper sought to improve the classification accuracy by reducing the error rate as observed from the conducted experiments on different data sets.

It is found that the mRMR features are less correlated with each other in comparison to the features extracted by the other methods.

REFERENCES

[1] Peng ,H.Long,F.Ding, C., Feature Selection Based on Mutual Information: Criteria of Max-Relevance,and Min-Redundancy,.In IEEE Transaction on Pattern Analysis and Machine Intelligence,vol(27),pp 1226-1238, 2005.

[2] J.O.Kephart and B. Arnold. A Feature Selection and Evaluation of Computer Virus Signatures. In Proc. Of the 4th Virus Bulletin International Conference, 1994, pp. 178-184.

[3] J.O.Kephart and B. Arnold, N-grams-Based File Signatures For Malware Detection, 1994,pp. 178-184

[4] Matthew G. Schultz, Eleazar Eskin, Erez Zadok, and Salvatore J.Stolfo, Data Mining Methods for detection of New Malicious Executables, In Proc. Of the IEEE Symposium on Security and Privacy,2001.

[5] Jeremy Z. Kolter and Marcus A.Maloof, Learning to Detect Malicious Executables in the Wild, In Proc. Of the tenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2004,pp 470-478.

[6] M. Stamp. Information Security: Principles and Practice, John Wiley and Sons,2011.

[7] Wong ,Wing and Stamp, Mark , Low Richard M and Stamp Mark.Hunting for metamorphic engines .Journal in Computer Virology,Vol.2:211-229,2006.

[8] Olivier Henchiri and Nathalie Japkowicz, A Feature Selection and Evaluation Scheme for Computer Virus Detection, In Proc. Of ICDM' 06,2006,pp. 891-895.

[9] C. Ding and H.C.Peng,"Minimum Redundancy Feature Selection from Microarray Gene Expression Data," Proc. Second IEEE Computational Systems Bioinformatics Conf.,pp. 523-528,Aug.2003.

[10] H. C. Peng and C.Ding,"Structural Search and Stability Enhancement of Bayesian Networks," Proc. Third IEEE Int'l Conf. Data Mining, pp.621-624, Nov.2003.

[11] A.Webb, Statistical Pattern Recognition. Arnold,1999.

[12] V. Vapnik, The Nature of Statistical Theory.New York :Springer 1995.

[13] T. Mitchell ,Machine Learning. McGraw-Hill, 1997.

[14] H.C.Peng , E.H. Herskovits and C.Davatzikos, "Bayesian Clustering Methods for Morphological Analysis of MR Images" Proc. Int'l Symp.Biomedical Imaging: from Nano to Macro, pp.485-488,2002.

[15] R .Kohavi and G. John, "Wrapper for Feature Subset Selection"Artifical Intelligence,vol. 97,nos.1-2,pp.273-324, 1997.

[16] A. Hyvarinen, J.Karhunen, and E.Oja, Independent Component Analysis. John Wiley and Sons,2001.