RSF: Roughset Theory Based Fuzzy Classification in Randomized Dimensionality Feature Selection

D. Barathi
Assistant professor, Department of Computer Science,
Dr.R.V.Arts and Science College, Karamadai, Coimbatore-641 104

Abstract

Feature Selection is a pattern dimensionality using feature mining and feature selection fit in to the data mining. To improve the robustness of the feature selection algorithm and for visualization point the dimension reduction techniques may be analyzed. The randomized feature selection is the transformation of high-dimensional data into animportantdesign of reduced dimensionality that corresponds to the fundamental dimensionality of the data. The normal reduction algorithm often not well for large datasets and fault dimensionality reduction, hence, to enhance the efficiency, the proposed system apply Roughset theory based fuzzy classification on original data set and obtain a reduced dataset containing possibly uncorrelated variables. In this paper, Roughset theory for feature selection and fuzzy based classification for Feature selection non-linear conversion is used for reduce the dimensionality and primary crisp value is calculated, then it is applied to classification algorithm.

Index Terms - *Rough set, Fuzzy, preprocessing.*

I. INTRODUCTION

Rough set theory can be regarded as a new mathematical tool for imperfect data analysis. The theory has found applications in many domains, such as decision support, engineering, environment, banking, medicine and others.

The fundamental concept behind Rough Set Theory is the approximation of lower and upper spaces of a set, the approximation of spaces being the formal classification of knowledge regarding the interest domain.

The subset generated by lower approximations is characterized by objects that will definitely form part of an interest subset, whereas the upper approximation is characterized by objects that will possibly form part of an interest subset. Every subset defined through upper and lower approximation is known as Rough Set

Feature selection, is a problem closely related to dimension reduction. The objective of feature selection is to identify features in the data-set as important, and discard any other feature as irrelevant and redundant information. Since feature selection reduces the dimensionality of the data, it holds out the possibility of more effective & rapid operation of data mining algorithm (i.e. Data Mining algorithms can be operated faster and more effectively by using feature selection).

There are often many features in KDD, and combinatorial large numbers of feature combinations, to select from. Note that the number of feature subset combinations with m features from a collection of N total features is $N!=[m!(N-m)!]$. It might be expected that the inclusion of an increasing number of features would increase the likelihood of including enough information to distinguish between classes. Unfortunately, this is not true if the size of the training dataset does not also increase rapidly with each additional feature included.

Rule-based expert systems are often applied to classification problems in various application fields, like fault detection, biology, and medicine. Fuzzy logic can improve such classification and decision support systems by using fuzzy sets to define overlapping class definitions. The application of fuzzy if-then rules also improves the interpretability of the results and provides more insight into the classifier structure and decision making process. To focus on the extraction of fuzzy rule-based classifiers from labeled data. Data-driven identification of such classifiers has to deal with structural issues, like the selection of the relevant features and finding an effective partitioning of the input domain. Moreover, linguistic interpretability is also an important aspect of rule-based classifier

The rest of this paper is organized as follows. In Section 2 review the existing related work. The proposed models and descriptions are described in Section 3. Finally conclude the paper in Section 4.

II. LITERATURE SURVEY

D. Aha and D. Kibler [1] illustrated a framework and methodology, called instance-based learning that

generates classification predictions using only specific instances. Instance-based learning algorithms do not maintain a set of abstractions derived from specific instances. This approach extends the nearest neighbor algorithm, which has large storage requirements. The authors described how storage requirements can be significantly reduced with, at most, minor sacrifices in learning rate and classification accuracy. While the storage-reducing algorithm performs well on several real world databases, its performance degrades rapidly with the level of attribute noise in training instances. Therefore, this paper extended it with a significance test to distinguish noisy instances. This extended algorithm's performance degrades gracefully with increasing noise levels and compares favorably with a noise-tolerant decision tree algorithm.

F. Alonso-Atienza et.al [2] proposed an early detection of ventricular fibrillation (VF) is crucial for the success of the defibrillation therapy in automatic devices. A high number of detectors have been proposed based on temporal, spectral, and time–frequency parameters extracted from the surface electrocardiogram (ECG), showing always a limited performance. The combination ECG parameters on different domain (time, frequency and time–frequency) using machine learning algorithms has been used to improve detection efficiency. However, the potential utilization of a wide number of parameters benefiting machine learning schemes has raised the need of efficient feature selection (FS) procedures. In this study, the authors proposed a novel FS algorithm based on support vector machines (SVM) classifiers and bootstrap resampling (BR) techniques. To define a backward FS procedure that relies on evaluating changes in SVM performance when removing features from the input space. This evaluation is achieved according to a nonparametric statistic based on BR. After simulation studies, author benchmarked the performance of his FS algorithm in AHA and MIT-BIH ECG databases. This paper results show that the proposed FS algorithm outperforms the recursive feature elimination method in synthetic examples, and that the VF detector performance improves with the reduced feature set.

D. Dernoncourt, B. Hanczar, and J. D. Zucker [3] discussed a feature selection is an important step when building a classifier on high dimensional data .As the number of observations is small the feature selection tends to be unstable. It is common that two feature subsets, obtained from different datasets but dealing with the same classification problem, do not overlap significantly. Although it is a crucial problem, few works have been done on the selection stability. The behavior of feature selection is analyzed in various conditions, not exclusively but with a focus on t-score based feature selection approaches and small sample data. The analysis is in three steps: the first one is theoretical using a simple mathematical model; the second one is empirical and based on artificial data; and the last one is based on real data. These three analyses lead to the same results and give a better understanding of the feature selection problem in high dimension data.

J. Fan and Y. Fan [4] proposed data classification using high-dimensional features arises frequently in many contemporary statistical studies such as tumor classification using microarray or other high-throughput data. The impact of dimensionality on classifications is largely poorly understood. To demonstrate that even for the independence classification rule, classification using all the features can be as bad as the random guessing due to noise accumulation in estimating population centroids in high-dimensional feature space. In fact, to demonstrate further that almost all linear discriminates can perform as bad as the random guessing. Thus, it is vital important to select a subset of important features for high dimensional classification, resulting in Features Annealed Independence Rules (FAIR). The conditions under which all the important features can be selected by the two-sample t-statistic are established. The choice of the optimal number of features, or equivalently, the threshold value of the test statistics are proposed based on an upper bound of the classification error. Simulation studies and real data analysis support our theoretical results and demonstrate convincingly the advantage of our new classification procedure.

A. J. Ferreira and M. A. T. Figu [5] discussed about a feature selection is a central problem in machine learning and pattern recognition. On large datasets (in terms of dimension and/or number of instances), using search-based or wrapper techniques can be computationally prohibitive. Moreover, many filter methods based on relevance/redundancy assessment also take a prohibitively long time on high-dimensional datasets. The authors proposed an efficient unsupervised and supervised feature selection/ranking filters for high-dimensional datasets. These methods use low-complexity relevance and redundancy criteria, applicable to supervised, semi-supervised, and unsupervised learning, being able to act as pre-processors for computationally intensive methods to focus their attention on smaller subsets of promising features. The experimental results, with up to 105 features, show the time efficiency of our methods, with lower generalization error than state-of-the-art techniques, while being dramatically simpler and faster.

Y. Han and L. Yu [6] presented a theoretical framework about the relationship between the stability and accuracy of feature selection based on a formal bias-variance decomposition of feature selection error. The framework also suggests a

variance reduction approach for improving the stability of feature selection algorithms. Furthermore, the author proposed an empirical variance reduction framework, margin based instance weighting, which weights training instances according to their influence to the estimation of feature relevance. In this paper author also developed an efficient algorithm under this framework. Experiments based on synthetic data and real-world microarray data verify both the theoretical framework and the effectiveness of the proposed algorithm on variance reduction. The proposed algorithm is also shown to be effective at improving subset stability, while maintaining comparable classification accuracy based on selected features.

J. Hua, W. D. Tembe, and E. R. Dougherty [7] studied a contemporary biological technologies produce extremely high-dimensional datasets from which to design classifiers, with 20,000 or more potential features being common place. In addition, sample sizes tend to be small. In such settings feature selection is an inevitable part of classifier design. Therefore, there have been a number of comparative studies for feature selection, but they have either considered settings with much smaller dimensionality than those occurring in current bioinformatics applications or constrained their study to a few real datasets. This study compared some basic feature-selection methods in settings involving thousands of features, using both model-based synthetic data and real data. It defines distribution models involving different numbers of markers (useful features) versus non-markers (useless features) and different kinds of relations among the features. Under this framework, it evaluates the performances of feature-selection algorithms for different distribution models and classifiers. Both classification error and the number of discovered markers are computed. Although the results clearly show that none of the considered feature-selection methods performs best across all scenarios, there are some general trends relative to sample size and relations among the features. For instance, the classifier-independent univariate filter methods have similar trends. Filter methods such as the t-test have better or similar performance with wrapper methods for harder problems. This improved performance is usually accompanied with significant peaking. Wrapper methods have better performance when the sample size is sufficiently large. Relief F, the classifier-independent multivariate filter method, has worse performance than univariate filter methods in most cases however, Relief F-based wrapper method show performance similar to their t-test-based counterparts.

L. Yu and H. Liu [8] proposed a feature selection is applied to reduce the number of features in many applications where data has Hundreds or Thousands of features. Existing feature selection methods mainly focus on finding relevant features. In this paper, the authors showed that feature relevance alone is insufficient for efficient feature selection of high-dimensional data. Authors defined a feature redundancy and proposes to perform explicit redundancy analysis in feature selection. A new framework is introduced that decouples relevance analysis and redundancy analysis. Authors developed a correlation-based method for relevance and redundancy analysis, and conduct an empirical study of its efficiency and effectiveness comparing with representative methods.

III. PROPOSED METHODOLOGY

A. Data Preprocessing

The data preprocessing of untrained raw dataset is first partitioned into three groups: (1) a predetermined set of instance initiation, (2) the group of attributes (features, variables) and (3) the class of attribute. For each groups in the dataset, a reduction decision classification is constructed. For each reduction system is consequentlydivided into two parts: the training dataset and the testing dataset. Each training dataset uses the corresponding input features and fall into two classes: normal (+1) and abnormal (−1).

B.Feature Selection

The Roughest based feature selection as the process of finding a subset of attributes, from the original raw dataset of pattern features, optimally according to the defined criterion. Rough sets theory is based on the concept of a lower and anupper approximation of a set, the approximation space and models of sets.

An information system can be represented as,

$$R = (U, A, V, f); \quad (1)$$

where U is the universe, a finite set of N objects $(x_1, x_2, \ldots, x_N)$ (a nonempty set), A is a finite set of attributes, $V = U_{a \in A} V_a$ (where V_a is a domain of the attribute a), $f : U \times A \rightarrow V$ is the total decision function (called the information function) such that $f(x, a) \in Va$ for every a $\in$A, $x \in U$. B subset of attributes $B \subseteq Q$ defines an equivalence relation (called an in-discernibility (unnoticeable) relation) on U.

$$\text{IND(A)} = \{(x,y) \in U : \text{for all a} \in B; f(x, a) = f(y, a)\}, \quad (2)$$

denoted also by A'. The information system can also bedefined as a decision table

$$DT = (U, C \cup D, V, f), \quad (3)$$

where C is a set of condition attributes, D is a set of decision attributes, $V = U_{a \in C \cup D} V_a$, where Va is the set of the domain of an attribute $a \in Q$, $f : U \times (C \cup D) \rightarrow V$ is a total decision function (information function, decision rule in DT) such that $f(x, a) \in Vq$ for every $a \in A$ and $x \in V$.

The straightforward feature selection procedures are based on an evaluation of the predictivepower of individual features, followed by a ranking of such evaluated features and eventually the choice of the first best m features. A criterion applied to an individual feature could be either of the open-loop or closed-loop type. It can be expected that a single feature alone may have a very low predictive power, whereas when put together with others, it may demonstrate a significant predictive power.

The final best features can be found by calculating the dependency determines between any uncertain feature and the decisional feature. After that, a ranking of features can be done. A basic filter algorithm to perform feature selection based on rough sets is shown in algorithm 1. This method calculates the dependency between every conditional feature considering the decisional feature, after ranking only the features with higher dependency values are included in the final subset of best features.

Algorithm1:Feature selection based on Rough sets

Input: Set of conditional and decisional features C, D.
Output: A subset of features
Process
Step 1: Initialize the best subset of features as the empty set.
Step 2: For i in 1: number of conditional features Apply some evaluation measure based on dependency of Rough sets.
End for
Step 3: Order the features according to dependency measure
Step 4: Select only the features with high dependency measure.

The lower approximation is a description of the feature objects which are known to belong to the concept of interest with full certainty, whilst the upper approximation is the set of all those objects which definitely or possibly belong to the concept of interest. The difference between the upper and lower approximation is a n area known as the boundary region or region of uncertainty.

C. Fuzzy-Roughest Classification

This classification approach new definitions of fuzzy based roughest lower and upper approximations are constructed that avoid the use of fuzzy logical connectives altogether. The logical decision rules are induced from lower and upper approximations defined for positive and negative relationships between credibility of premises and conclusions.

The Feature selection reads a set of feature patterns and outputs another with reduced dimensionality, implemented with the Roughset algorithm. The rule induction reads a sequence of feature patterns and outputs a set of if-then rules connecting features and their implied classes. The Fuzzy classification a standard approximate reasoner that interprets the induced fuzzy rule-set and uses it to classify previously unseen feature pattern.

The fuzzy classification rules that each describe one of the N_c classes in the data set. The rule antecedent is a fuzzy description in the n-dimensional feature space and the rule consequent is a crisp (non-fuzzy) class label from the set $\{1, 2, \ldots, N_c\}$:

IV. CONCLUSION AND FUTURE WORK

In this paper proposed a novel Roughset feature selection based fuzzy classification to enforce attribute selection algorithm. Theproposed processing sequence has shown a possible for feasible feature extraction and feature selection in designing of fuzzy based classifiers for real world datasets. Meanwhile proposed method provides significant reduction of pattern dimensionality. Rough set methods have shown capability to reduce significantly the pattern dimensionality and have proven to be viable data mining techniques as a front end of neural network classifiers.

In future work, we intend to enhance the proposed methodology to develop the some other classification based scheme of the result data.

REFERENCES

[1] D. Aha and D. Kibler, “Instance-based learning algorithms,” Mach. Learn., vol. 6, no. 1, pp. 37–66, 1991.
[2] F. Alonso-Atienza, J. L. Rojo-Alvare, A. Rosado-Mu~noz, J. J. Vinagre, A. Garcia-Alberola, and G. Camps-Valls, “Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection,” Expert Syst. Appl., vol. 39, no. 2, pp. 1956–1967, 2012.
[3] D. Dernoncourt, B. Hanczar, and J. D. Zucker, “Analysis of feature selection stability on high dimension and small sample data,” Comput. Statist. Data Anal., vol. 71, pp. 681–693, 2014.
[4] J. Fan and Y. Fan, “High dimensional classification using features annealed independence rules,” Ann. Statist., vol. 36, no. 6, pp. 2605–2637, 2008.
[5] A. J. Ferreira and M. A. T. Figueiredo, “Efficient feature selection filters for high dimensional data,” Pattern Recog. Lett., vol. 33, no. 13, pp. 1794–1804, 2012.
[6] Y. Han and L. Yu, “A variance reduction framework for stable feature selection,” Statist. Anal. Data Mining, vol. 5, no. 5, pp. 428–445, 2012.
[7] J. Hua, W. D. Tembe, and E. R. Dougherty, “Performance of feature-selection methods in the classification of high-dimension data,” Pattern Recognit., vol. 42, no. 3, pp. 409–424, 2009.
[8] L. Yu and H. Liu, “Efficient feature selection via analysis of relevance and redundancy,” The J. Mach. Learn. Res., vol. 5, no. 2, pp. 1205–1224, 2004.