International Journal of Research in Advent Technology, Vol.2, No.9, September 2014
E-ISSN: 2321-9637

Survey on Rough Set Feature Selection Using Evolutionary Algorithm

M.Gayathri[1], Dr.C.Yamini[2]
Research Scholar[1], Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore [1]
Assistant Professor [2], Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore [2]
Email: m.gayathrinathan@gmail.com[1] , c_yamini@yahoo.com[2]

Abstract- Most of the conventional feature selection algorithms have a drawback whereby a weakly ranked gene that could perform well in terms of classification accuracy with an appropriate subset of genes will be left out of the selection. Considering this shortcoming, various authors proposed a feature selection algorithm in gene expression data analysis of sample classifications. Feature selection has been the focus of interest for quite some time and much work has been done. With the creation of huge databases and the consequent requirements for good machine learning techniques, new problems arise and novel approaches to feature selection are in demand]. This method shows promising classification accuracy for all the test data sets. They also show the relevance of the selected genes in terms of their biological functions.

Index Terms- *Ranked gene, Classification Accuracy, Machine Learning.*

1. INTRODUCTION

Feature selection, also known as subset selection or variable selection, is a process commonly used in machine learning, wherein a subset of the features available from the data is selected for application of a learning algorithm. Feature selection is necessary either because it is computationally infeasible to use all available features, or because of problems of estimation when limited data samples (but a large number of features) are present. The latter problem is related to the so-called curse of dimensionality. In machine learning and statistics, feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. The central assumption when using a feature selection technique is that the data contains many *redundant* or *irrelevant* features.

2. A DISTANCE MEASURE APPROACH TO EXPLORING THE ROUGH SET BOUNDARY REGION FOR ATTRIBUTE REDUCTION

Feature selection is a process which chooses a subset of the original features present in a given data set which provides the most useful information. Following selection, the most important information of the data set should still remain. In fact, efficient FS techniques should be able to detect and ignore noisy and misleading features. As a result, the data set quality may even increase through feature selection. Classification accuracy may be increased as a result of feature selection through the removal of noisy, irrelevant, or redundant features. Also in domains where features correspond to measurements, fewer features offer advantages such as minimizing the expense and time consumed in recording such measurements.

The work on rough set theory (RST) offers a formal methodology that can be employed to reduce the dimensionality of data sets, as a preprocessing step to assist any chosen modeling method for learning from data. It assists in identifying and selecting the most information-rich features in a data set. This is achieved without transforming the data, while simultaneously attempting to minimize information loss during the selection process. In terms of computational effort, this approach is highly efficient, and is based on simple set operations, which makes it suitable as a preprocessor for techniques that are much more complex. In contrast to statistical correlation-reduction approaches, RST requires no human input or domain knowledge other than the given data sets. Perhaps most importantly though, it retains the underlying semantics of the data, which results in data models that are more transparent to human scrutiny.

Many approaches based on rough set theory up to now, have employed the dependency function, which is based on lower approximations as an evaluation step in the FS process. However, by examining only that information which is considered to be certain and ignoring the boundary region, or region of uncertainty, much useful information is lost. This work examines a rough set FS technique which uses the information gathered from both the lower approximation dependency value and a distance metric which considers the number of objects in the boundary region and the distance of those objects from the lower approximation. The use of this measure in rough set feature selection can result in smaller subset sizes than those obtained using the dependency function alone. This demonstrates that there is much valuable information to be extracted from the boundary region.

3. ON THE GENERALIZATION OF FUZZY ROUGH SETS

The concept of rough set was originally proposed as a mathematical approach to handle imprecision, vagueness, and uncertainty in data analysis. This theory has amply been demonstrated to have its usefulness and versatility by successful applications in a variety of problems. The theory of rough sets deals with the approximation of an arbitrary subset of a universe by two definable or observable subsets called lower and upper approximations. By using the concepts of lower and upper approximations in rough set theory, knowledge hidden in information systems may be unraveled and expressed in the form of decision rules. Another particular use of rough set theory is that of attribute reduction in databases. Given a dataset with discretized attribute values, it is possible to find a subset of the original attributes that are the most informative. This leads to the concept of attributes reduction which can be viewed as the strongest and most characteristic results in rough set theory to distinguish itself from other theories.

Fuzzy rough sets encapsulate the related but distinct concepts of fuzziness and indiscernibility, both of which occur as a result of uncertainty in knowledge or data, thus a method employing fuzzy rough sets should be adopted to handle this uncertainty. There are at least two approaches for the development of the fuzzy rough set theory, the constructive and axiomatic approaches. In constructive approach, fuzzy relations on the universe is the primitive notion, the lower and upper approximation operators are constructed by means of this notion.

Rough sets and fuzzy sets have been proved to be powerful mathematical tools to deal with uncertainty; it soon raises a natural question of whether it is possible to connect rough sets and fuzzy sets. The existing generalizations of fuzzy rough sets are all based on special fuzzy relations (fuzzy similarity relations, -similarity relations), it is advantageous to generalize the fuzzy rough sets by means of arbitrary fuzzy relations and present a general framework for the study of fuzzy rough sets by using both constructive and axiomatic approaches. In this work, from the viewpoint of constructive approach, we first propose some definitions of upper and lower approximation operators of fuzzy sets by means of arbitrary fuzzy relations and study the relations among them, the connections between special fuzzy relations and upper and lower approximation operators of fuzzy sets are also examined. In axiomatic approach, here characterize different classes of generalized upper and lower approximation operators of fuzzy sets by different sets of axioms. The lattice and topological structures of fuzzy rough sets are also proposed.

4. SEMANTICS-PRESERVING DIMENSIONALITY REDUCTION: ROUGH AND FUZZY-ROUGH-BASED APPROACHES

Many problems in machine learning involve high dimensional descriptions of input features. It is therefore not surprising that much research has been carried out on dimensionality reduction. However, existing work tends to destroy the underlying semantics of the features after reduction or require additional information about the given data set for thresholding. A technique that can reduce dimensionality using information contained within the data set and that preserves the meaning of the features (i.e., semantics-preserving) is clearly desirable. Rough set theory (RST) can be used as such a tool to discover data dependencies and to reduce the number of attributes contained in a data set using the data alone and no additional information.

It is, therefore, desirable to develop techniques to provide the means of data reduction for crisp and real value attributed data sets which utilizes the extent to which values are similar. This can be achieved through the use of fuzzy-rough sets. Fuzzy-rough sets encapsulate the related but distinct concepts of vagueness and indiscernibility (for rough sets), both of which occur as a result of uncertainty in knowledge.

Semantics-preserving dimensionality reduction refers to the problem of selecting those input features that are most predictive of a given outcome; a problem encountered in many areas such as machine learning, pattern recognition, and signal processing. This has found successful application in tasks that involve data sets containing huge numbers of features (in the order of tens of thousands), which would be impossible to process further. Recent examples include text processing and Web content classification. One of the many successful applications of rough set theory has been to this feature selection area. This work reviews those techniques that preserve the underlying semantics of the data, using crisp and fuzzy rough set-based methodologies.

5. FUZZY PROBABILISTIC APPROXIMATION SPACES AND THEIR INFORMATION MEASURES

Rough set methodology has been witnessed great success in modeling with imprecise and incomplete information. The basic idea of this method hinges on classifying objects of discourse into classes containing indiscernible objects with respect to some attributes. Then the indiscernible classes, also called knowledge granules, are used to approximate the unseen object sets. In this framework, an attribute set is viewed as a family of knowledge, which partitions the universe into some knowledge granules or elemental concepts. Rough set theory has proven to be an efficient tool for modeling and reasoning with uncertainty information. By introducing probability into fuzzy approximation space, a theory about fuzzy probabilistic approximation spaces is proposed in this work, which combines three types of uncertainty: probability, fuzziness, and roughness into a rough set model. Here introduce Shannon's entropy to measure information quantity implied in a Pawlak's approximation space, and then present a novel representation of Shannon's entropy with a relation matrix. Based on the modified formulas, some generalizations of the entropy are proposed to calculate the information in a fuzzy approximation space and a fuzzy probabilistic approximation space, respectively.

In the rough set framework, attributes are called knowledge, which is used to form a concept system of the universe. Knowledge introduced by an attribute set implies in the partition of a referential universe. The more knowledge there is, the finer partition will be, and correspondingly it get a more perfect approximation of a subset in the universe. Attributes induce an order or a structure of universe of discourse, which decreases uncertainty or chaos of the universe. Given a universe, a probability distribution on, and some nominal, real-value or fuzzy attributes, there comes forth an interesting problem: How do measure the knowledge quantity introduced by an attribute set in the approximation space. In other words, it's interesting in constructing a measure to compute the discernibility power induced by a family of attributes. This measure leads to the likelihood to compare the knowledge quantity formed by different attributes, and help to find the important attribute set and redundancy of an information system.

In this work, Shannon's entropy is first introduced to compute the knowledge quantity of nominal attributes in Pawlak's approximation space, and then an extended information measure will be presented, which is suitable for the spaces where fuzzy attributes or fuzzy relations are defined on. Based on the extension, the solutions to measuring the information in fuzzy and fuzzy probabilistic hybrid approximation spaces are presented.

6. FEATURE SELECTION USING F-INFORMATION MEASURES IN FUZZY APPROXIMATION SPACES

Feature selection or dimensionality reduction of a data set is an essential preprocessing step used for pattern recognition, data mining, machine learning, etc. It is an important problem related to mining large data sets, both in dimension and size. Prior to analysis of the data set, preprocessing the data to obtain a smaller set of representative features and retaining the optimal salient characteristics of the data not only decrease the processing time, but also lead to more compactness of the models learned and better generalization. Rough set theory is a new paradigm to deal with uncertainty, vagueness, and incompleteness. It has been applied to fuzzy rule extraction, reasoning with uncertainty, fuzzy modeling, classification, feature selection, etc.. However, there are usually real-valued data and fuzzy information in real-world applications. Combining fuzzy and rough sets provides an important direction in reasoning with uncertainty for real-valued data sets. Both fuzzy and rough sets provide a mathematical framework to capture uncertainties associated with the data.

In this work, a novel feature selection method is proposed, which employs fuzzy-rough sets to provide a means by which discrete- or real-valued noisy data (or a mixture of both) can be effectively reduced without the need for user-specified information. Moreover, the proposed method can be applied to data with continuous or nominal decision attributes, and can be applied to regression as well as classification data sets. The proposed method selects a subset of features from the whole feature set by maximizing the relevance and minimizing the redundancy of the selected features. The relevance and redundancy of the features are calculated using the f-information measures in fuzzy approximation spaces. Using the concept of fuzzy equivalence partition matrix, the f-information measures are calculated for both condition and decision attributes. Hence, the only information required in the proposed feature selection method is in the form of fuzzy partitions for each attribute, which can be automatically derived from the given data set. Several quantitative measures are introduced based on fuzzy-rough sets to evaluate the performance of the proposed feature selection method. The effectiveness of the proposed method, along with a comparison with other methods, is demonstrated on a set of real-life data.

7. ATTRIBUTES REDUCTION USING FUZZY ROUGH SETS

The concept of rough sets was originally proposed by Pawlak as a mathematical approach to handling imprecision, vagueness and uncertainty in data analysis. This theory has amply been demonstrated to have its usefulness and versatility in successfully solving a variety of problems. One important application of rough sets theory is that of attributes reduction in databases. Given a dataset with discretized attribute values, it is possible to find a subset of the original attributes that contains the same information as the original one. The concept of attributes reduction can be viewed as the strongest and most important results in rough sets theory to distinguish itself from other theories. Rough sets approach of attributes reduction can be used as a purely structural method for reducing dimensionality using information contained within the dataset and preserving the meaning of the features.

The existing researches on fuzzy rough sets are mainly concentrated on the approximations of fuzzy sets. These researches have been studied and discussed completely. In a pioneering work on attributes reduction with fuzzy rough sets is proposed. Formal concepts of fuzzy-rough attributes reduction were introduced and an algorithm to compute a reduction was developed by using the dependence function. This algorithm had been tested with some practical data sets such as web categorization and was claimed to perform well. However, there are several aspects in their algorithm that are argumentative.

In this work, the concept of attributes reduction with fuzzy rough sets is proposed after a detailed analysis of the algorithm. The structure of reduction is completely studied and an algorithm using discernibility matrix to compute all the attributes reductions is developed. Thus a solid mathematical foundation is set up for attributes reduction with fuzzy rough sets. This work mainly focuses on the attributes reduction with fuzzy rough sets. After analyzing the previous works on attributes reduction with fuzzy rough sets, here introduce formal concepts of attributes reduction with fuzzy rough sets and completely study the structure of attributes reduction. An algorithm using discernibility matrix to compute all the attributes reductions is developed. Based on these lines of thought, here set up a solid mathematical foundation for attributes reduction with fuzzy rough sets.

8. NEIGHBORHOOD ROUGH SET BASED HETEROGENEOUS FEATURE SUBSET SELECTION

Feature subset selection as a common technique used in data preprocessing for pattern recognition, machine learning and data mining, has attracted much attention in recent years. Due to the development of information acquirement and storage, tens, hundreds, or even thousands of features are acquired and stored in databases for some real-world applications. With a limited amount of training data, an excessive amount of features may cause a significant slowdown in the learning process, and may increase the risk of the learned classifier to over-fit the training data because irrelevant or redundant features confuse learning algorithms. It is desirable to reduce data to get a smaller set of informative features for decreasing the cost in measuring, storing and transmitting data, shortening the process time and leading to more compact classification models with better generalization. The representative is attribute reduction based on rough set theory while the latter views all attributes as real-valued variables, which take values in the real-number spaces.

Feature subset selection is viewed as an important preprocessing step for pattern recognition, machine learning and data mining. Most of researches are focused on dealing with homogeneous feature selection, namely, numerical or categorical features. In this work, it introduce a neighborhood rough set model to deal with the problem of heterogeneous feature subset selection. As the classical rough set model can just be used to evaluate categorical features, here generalize this model with neighborhood relations and introduce a neighborhood rough set model. The proposed model will degrade to the classical one if its specify the size of neighborhood zero. The neighborhood model is used to reduce numerical and categorical features by assigning different thresholds for different kinds of attributes. In this model the sizes of the neighborhood lower and upper approximations of decisions reflect the discriminating capability of feature subsets. The size of lower approximation is computed as the dependency between decision and condition attributes. it use the neighborhood dependency to evaluate the significance of a subset of heterogeneous features and construct forward feature subset selection algorithms. The proposed algorithms are compared with some classical techniques. The main contributions of the work are two-fold. First, here extend the neighborhood rough set model to deal with data with heterogeneous features and discuss two classes of monotonicity in terms of consistency, neighborhood sizes and attributes; second, two efficient algorithms are designed for searching an effective feature subset.

9. FAST: A ROC-BASED FEATURE SELECTION METRIC FOR SMALL SAMPLES AND IMBALANCED DATA CLASSIFICATION PROBLEMS

One of the greatest challenges in machine learning and data mining research is the class imbalance problem presented in real world applications. The class imbalance problem refers to the issues that occur when a dataset is dominated by a class or classes that have significantly more samples that the other classes of the dataset. Imbalanced classes are seen in a variety of domains and many have major economic, commercial, and environmental concerns. Some examples include text classification, risk management, web categorization, medical diagnosis/monitoring, biological data analysis, credit card fraud detection, oil spill identification from satellite images. While the majority of learning methods are designed for well balanced training data, data imbalance presents a unique challenging problem to classifier design when the misclassification costs for the two classes are different (i.e., cost sensitive classification) and accordingly, the overall classification rate is not appropriate to evaluate the performance. The class imbalance problem could hinder the performance of standard machine learning methods. The sampling techniques and algorithmic methods may not work well for high dimensional class imbalance problems. Indeed, van der Putten and van Someren analyzed the COIL challenge 2000 datasets and concluded that to overcome over fitting problems, feature selection is even more important than classification algorithms. A similar observation was made by Forman in highly imbalanced data classification problems.

The class imbalance problem is encountered in a large number of practical applications of machine learning and data mining, for example, information retrieval and filtering, and the detection of credit card fraud. It has been widely realized that this imbalance raises issues that are either nonexistent or less severe compared to balanced class cases and often results in a classifier's suboptimal performance. This is even more true when the imbalanced data are also high dimensional. In such cases, feature selection methods are critical to achieve optimal performance. In this work, here propose a new feature selection method, Feature Assessment by Sliding Thresholds (FAST), which is based on the area under a ROC curve generated by moving the decision boundary of a single feature classifier with thresholds placed using an even-bin distribution. FAST is compared to two commonly-used feature selection methods, correlation coefficient and RELevance In Estimating Features (RELIEF), for imbalanced data classification.

10. MARGIN BASED FEATURE SELECTION - THEORY AND ALGORITHMS

Feature selection is closely related to the more general problems of dimensionality reduction and efficient data representation. Many dimensionality reduction methods, like Principal Component Analysis or Locally Linear Embedding are in fact unsupervised feature extraction algorithms, where the obtained lower dimensions are not necessarily subsets of the original coordinates. Other methods, more related to supervised feature extraction, are the Information Bottleneck and Sufficient Dimensionality Reduction. However, on many cases, feature selection algorithms provide a much simpler approach as they do not require the evaluation of new complex functions of the irrelevant features.

In the filter model the selection is done as a preprocessing, without trying to optimize the performance of any specific predictor directly. This is usually achieved through an (ad-hoc) evaluation function using a search method in order to select a set that maximizes this function. Performing an exhaustive search is usually intractable due to the large number of initial features. Different methods apply a variety of search heuristics, such as hill climbing and genetic algorithms. One commonly used evaluation function is the mutual information between the feature set and the labels.

Feature selection is the task of choosing a small set out of a given set of features that capture the relevant properties of the data. In the context of supervised classification problems the relevance is determined by the given labels on the training data. A good choice of features is a key for building compact and accurate classifiers. In this work, here introduce a margin based feature selection criterion and apply it to measure the quality of sets of features. Using margins it devise novel selection algorithms for multi-class classification problems and provide theoretical generalization bound. here also study the well known Relief algorithm and show that it resembles a gradient ascent over our margin criterion. Here apply our new algorithm to various datasets and show that the new Simba algorithm, which directly optimizes the margin, outperforms Relief.

11. FEATURE SELECTION BASED ON MUTUAL INFORMATION: CRITERIA OF MAX-DEPENDENCY

Feature selection is an important problem for pattern classification systems. here study how to select good features according to the maximal statistical dependency criterion based on mutual information. Because of the difficulty in directly implementing the maximal dependency condition, the first derive an equivalent form, called minimal-redundancy-maximal-relevance criterion (mRMR), for first-order incremental feature selection. Then, in this paper present a two-stage feature selection algorithm by combining mRMR and other more sophisticated feature selectors (e.g., wrappers). This allows to select a compact set of superior features at very low cost. In this paper perform extensive experimental comparison of our algorithm and other methods using three different classifiers (naive Bayes, support vector machine, and linear discriminate analysis) and four different data sets (handwritten digits, arrhythmia, NCI cancer cell lines, and lymphoma tissues). The results confirm that mRMR leads to promising improvement on feature selection and classification accuracy. Feature selection is an important problem for pattern classification systems. In this paper study how to select good features according to the maximal statistical dependency criterion based on mutual information. Because of the difficulty in directly implementing the maximal dependency condition, the first derive an equivalent form, called minimal-redundancy-maximal-relevance criterion (mRMR), for first-order incremental feature selection. Then, here present a two-stage feature selection algorithm by combining mRMR and other more sophisticated feature selectors (e.g., wrappers). This allows to select a compact set of superior features at very low cost. In this work perform extensive experimental comparison of our algorithm and other methods using three different classifiers (naive Bayes, support vector machine, and linear discriminate analysis) and four different data sets (handwritten digits, arrhythmia, NCI cancer cell lines, and lymphoma tissues). The results confirm that mRMR leads to promising improvement on feature selection and classification accuracy.

12. CONCLUSION

The present wok proposes Genetic algorithm for selecting the features. It is of great of importance to remove the noisy and irrelevant features and data samples embedded in data sets before applying some data mining techniques to analyze the data sets. The presented genetic algorithm dedicated for a particular feature selection problem encountered in genetic analysis of different datasets. The specification of this problem is not looking for single feature but for several associations of features that may be involved in the studied datasets. Results are promising for all types of datasets as the algorithm seems to be robust and to be able to isolate interesting associations.

REFERENCES

[1] N. Parthalain, Q. Shen, and R. Jensen, “A Distance Measure Approach to Exploring the Rough Set Boundary Region for Attribute Reduction,” IEEE Trans. Knowledge and Data Eng., vol. 22, no. 3, pp. 305-317, Mar. 2010.

[2] D.S. Yeung, D. Chen, E.C.C. Tsang, J.W.T. Lee, and W. Xizhao, “On the Generalization of Fuzzy Rough Sets,” IEEE Trans. Fuzzy Systems, vol. 13, no. 3, pp. 343-361, June 2005

[3] R. Jensen and Q. Shen, “Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approach,” IEEE Trans. Knowledge and Data Eng., vol. 16, no. 12, pp. 1457-1471, Dec. 2004.

[4] Q. Hu, D. Yu, Z. Xie, and J. Liu, “Fuzzy Probabilistic Approximation Spaces and Their Information Measures,” IEEE Trans. Fuzzy Systems, vol. 14, no. 2, pp. 191-201, Dec. 2007.

[5] P. Maji and S.K. Pal, “Feature Selection Using f-Information Measures in Fuzzy Approximation Spaces,” IEEE Trans. Knowledge and Data Eng., vol. 22, no. 6, pp. 854-867, June 2010.

[6] E.C.C. Tsang, D. Chen, D.S. Yeung, X.-Z. Wang, and J. Lee, “Attributes Reduction Using Fuzzy Rough Sets,” IEEE Trans. Fuzzy Systems, vol. 16, no. 5, pp. 1130-1141, Oct. 2008

[7] Q. Hu, D. Yu, J. Liu, and C. Wu, “Neighborhood Rough Set Based Heterogeneous Feature Subset Selection,” Information Sciences, vol. 178, pp. 3577-3594, 2008

[8] X.-W. Chen and M. Wasikowski, “FAST: A ROC-Based Feature Selection Metric for Small Samples and Imbalanced Data Classification Problems,” Proc. 14th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining, 2008.

[9] R. Gilad-Bachrach, A. Navot, and N. Tishby, “Margin Based Feature Selection: Theory and Algorithms,” Proc. 21st Int’l Conf. Machine Learning, 2004.

[10] H. Peng, F. Long, and C. Ding, “Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226-1238, Aug. 2005