



WWJMRD 2018; 4(1): 159-159
www.wwjmr.com
International Journal
Peer Reviewed Journal
Refereed Journal
Indexed Journal
UGC Approved Journal
Impact Factor MJIF: 4.25
e-ISSN: 2454-6615

D. Richard

Bharathiar University,
Coimbatore-, Tamil Nadu,
India

R. Mala

Dept of Computer Science,
Alagappa University, College
of Arts & Science, Parmakudi,
Tamil Nadu, India

Gr-Anfis Data Mining Approach For Envisage Heart Attack Disease

D. Richard, R. Mala

Abstract

Heart disease is the most important reason of death in the world over the past 20 years. Researchers have been using a number of data mining techniques to help health care professionals in the analysis of heart disease. Filter selection used for data reduction can be classified into three types are wrapper, filter and embedded method. These filter approach is independent of the learning algorithm, computationally simple fast and scalable. GR-ANFIS is one of the successful data mining techniques used and most research has useful J4.8 decision tree, based on Gain Ratio.

Keywords: Feature selection -filter approach, Gain Ratio, GR-ANFIS, J4.8 decision tree

1. Introduction

The data mining knowledge form large of data and cleaning to remove noise and inconsistent data, integration where multiple data sources may be combined to selection where data relevant to the analysis task are retrieved from the database of transformation where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance Data mining an essential process where intelligent methods are applied in order to extract data patterns Pattern evaluation to identify the truly interesting patterns representing knowledge based on some interestingness measures [1]. The feature selection is known as variable and selection, attribute selection or variable subset selection to process of selecting a subset of relevant features from the feature space for use in model construction and reducing the features from the feature space to a manageable size for processing and analysis [2].

The filter approach to independent of the classifier and Information, correlation, distance between inter and intra class another method wrapper detects to error rate used as a measure Sub set feature selection, performance good for particular type model only and another method is hybrid method to takes-all group of techniques which perform feature selection part of the model construction process [2]. In decision tree learning, Gain ratio is a information gain to the built-in information. It is used to reduce a bias towards multi-valued attributes by taking the number and size of branches into account when choosing an attribute. Information gain ratio biases the decision tree against considering attributes with a large number of separate values [3].

Classification is the process of building a model of classes from a set of records that contain class labels. Decision Tree Algorithm is to find out the way the attributes-vector behaves for a number of instances. Also on the bases of the training instances the classes for the newly generated instances are being found. This algorithm generates the rules for the prediction of the target variable. With the help of tree classification algorithm the critical distribution of the data is easily understandable [4]. A data set is a collection of related, discrete items of related data that may be accessed individually or in combination or managed as a whole entity. The data set collected for reputed hospital in various place. In this research from the approach were so promising. If further attempts are engaged in the application of Information Technology in diagnosing various diseases such as heart then efficient, timely and decent healthcare services will be realized.

Correspondence:

D. Richard

Bharathiar University,
Coimbatore-, Tamil Nadu,
India

2. Methodology

The learning and classification steps of decision induction algorithm are simple and fast. In general, induction algorithm classifiers have good accuracy and induction algorithms have been used for classification in many application areas, such as medicine.

2.1 Pre-Process Method Diagram

The below diagram the first steps to be process on input features to set all the second step process on features sub section in this process evaluate data to pass the induction

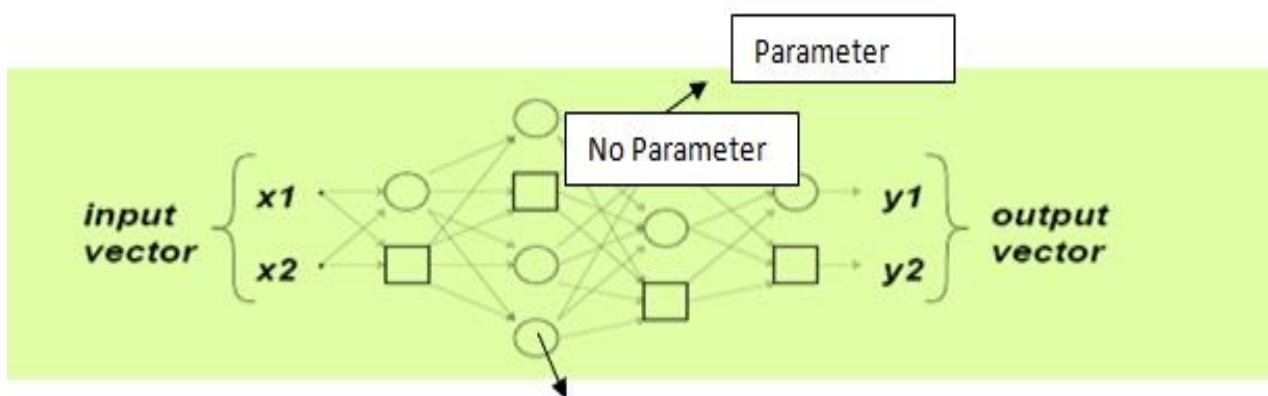


Features sub selection there are three type's wrapper, filter and hybrid. In this research to analysis to be take filter selection method. Because in this method is also called variable selection or attribute selection. It is the automatic selection of attributes in data (such as columns in tabular data) that are most relevant to the predictive modelling problem is work. Feature selection is different from dimensionality reduction.

algorithm, this algorithm to adopt J4.8 decision tree method, this method classify the data in proper way. The input features means to given data set inputs that is attributes, the attributes is age, sex, occupation, chest pain type (4 values), food habits, resting blood pressure, fasting blood sugar > 120 mg/dl, resting electrocardiographic results (values 0,1,2), maximum heart rate achieved, old peak = ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, thal:3=normal; 6=fixed defect; 7=reversible defect as follows 14 attributes as to be given input.

2.2 Anfis Data Flow and Process Diagram

An adaptive network is a multi-layer feed forward network in which each node performs a particular function on the incoming signals. The nature and the choice of the node function depend on the overall input-output function. No weights are associated with links and the links just indicate the flow. To achieve desired i/p-o/p mapping the parameters are updated according to training data and gradient-based learning procedure



3. GR-ANFIS Approach

A decision tree is a simple structure where non-terminal nodes represent tests on one or more attributes and terminal nodes reflect decision outcomes. It applies a kind of normalization to GR-ANFIS using a "split information" value defined analogously with *Info (D)* as,

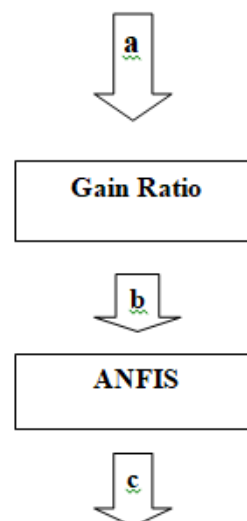
$$\text{Gain Ratio (A)} = \frac{\text{Gain (A)}}{\text{Split info (A)}}$$

The attribute with the maximum gain ratio is selected as the splitting attribute. Note, however, that as the split information approaches 0, the ratio becomes unstable. A constraint is added to avoid this, whereby the information gain of the test selected must be large at least as great as the average gain over all tests examined.

3.1 Proposed Structure of Gr-Anfis

The selected features were applied to ANFIS to train and test the proposed approach. The structure of the proposed approach where a= {a1, a2,..., an} are the original features in dataset, b={b1, b2,...,bs} are the features after applying

the Gain Ratio (split up values), and c denote the final output after applying b on ANFIS The Gain Ration is then fed to ANFIS



3.3 Steps of the Proposed Algorithm

- Step1:** Load data.
- Step2:** Generate the dataset
- Step3:** Now generate with given dataset.
- Step4:** Now using testing dataset with Information, correlation, and distance between inter and intra class and dummy values in it evaluate the file (dataset) using filter search method.
- Step5:** To remove the dummy values entries all the output values which has <0.4 value is removed i.e. not considered as proper data.
- Step6:** Now using the evaluated data, again J4.8 decision tree is generated.
- Step7:** Find out the GR-ANFIS ratio or percentage (%).
- Step8:** Compare this generated to result for (GR-ANFIS) with (GI-ANFIS) and (IG-ANFIS) using dataset.

4. Results and Analysis

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily [2, 7]. The discussion can be made in several sub-chapters.

4.1 NNS Algorithm

Nearest neighbour search (NNS), also known as proximity search, is the optimization problem of finding the point in a given set that is closest (or most similar) to a given point. Closeness is typically expressed in terms of a dissimilarity function: the less similar the objects, the larger the function values [6]. Formally, the nearest-neighbour (NN) search problem is defined as follows: given a set *S* of points in a space *M* and a query point *q* ∈ *M*, find the closest point in *S* to *q*. There are two disadvantage using NNs algorithm in general they can learn correctly from examples but what is learn is not easy for human to understand that is knowledge base extracted from them does not have such intuitive representation as that provide secondly the type of functions that can be used in NNs must process precise regularity features and the derivative of these functions have to be taken in this kind of algorithm.

4.2 ILFN-Fuzzy Algorithm

The ILFN network was developed for pattern classification applications. The ILFN network, which employed fuzzy sets and neural network theory, equips with a fast, one-pass, on-line, and incremental learning algorithm [7]. After trained, the ILFN network stored numerical knowledge in hidden units, which can then be directly interpreted into if then rule bases. The major draw focuses on rules of combining membership functions discussed above are known as the mini max rule for conjunctive (AND) and

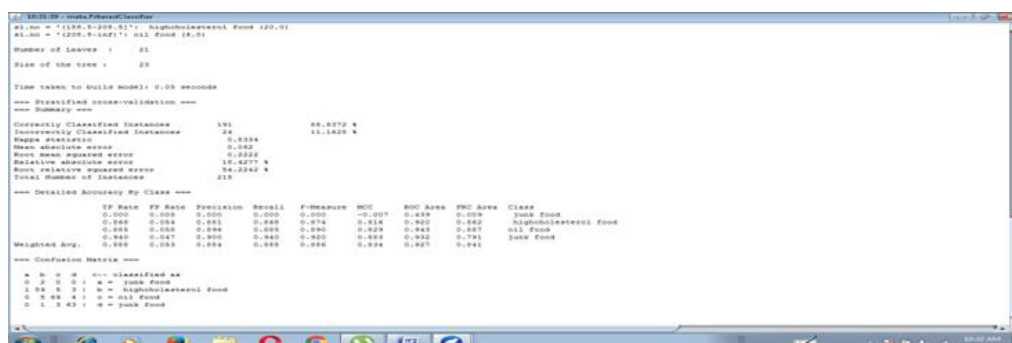
disjunctive (OR) reasoning. These rules have a major drawback: They are not robust at all. If we try to imitate the way humans' reason, the mini max rule is definitely not the way. Many researchers have proposed different rules of combining conjunctive or disjunctive clauses: for example, instead of taking the minimum or the maximum of the membership functions, they take the arithmetic or the geometric mean. These rules are arbitrary, and there are lots of them. It is possible, if we have enough training data, i.e. conditions and class assignments by the experts, to train our system so that it chooses the best rule that fits the way of reasoning of the expert that did the classification. Another disadvantage of the rules discussed earlier is that they give the same importance to all factors that are to be combined. For example, it is possible that the role of soil depth or rock permeability is not of the same importance to soil erosion as the role of slope. This issue can be resolved if we do not insist on all membership functions taking values between 0 and 1.

4.3 IG-ANFIS

IG is a feature selection technique used to reduce the number of input to ANFIS. It uses ranking method and is often used in text categorization [8]. The information gain is usually a not good measure for deciding the relevance of an attribute, it is not perfect. A notable problem occurs when information gain is applied to attributes that can take on a large number of distinct values. For example, suppose that one is building a decision tree for some data describing the customers of a business. Information gain is often used to decide which of the attributes are the most relevant, so they can be tested near the root of the tree. One of the input attributes might be the customer's credit card number. This attribute has high mutual information, because it uniquely identifies each customer, but we do not want to include it in the decision tree this biases the decision tree against considering attributes with a large number of distinct values. However, attributes with very low information values then appeared to receive an unfair advantage.

4.4 GI-ANFIS

The Classification and Regression Trees (CART), which described the generation of binary decision trees CART were invented independently of one another at around the same time, yet follow a similar approach for learning decision trees from training tuple. But one or more drawbacks GI-ANFIS are any index is that of oversimplification. To begin with, absolute and relative levels of inequality are difficult to measure source. The remove the dummy and duplicated data entries all the output values which has <0.4 value is removed.



To view of Filter classifier using 14 attributes and view of summary of filter classifier result, accuracy by class and confusion matrix. The dataset used for J4.8 is shown in

below result. It is same as the original dataset are using more than 200 values and 14 attributes.

```

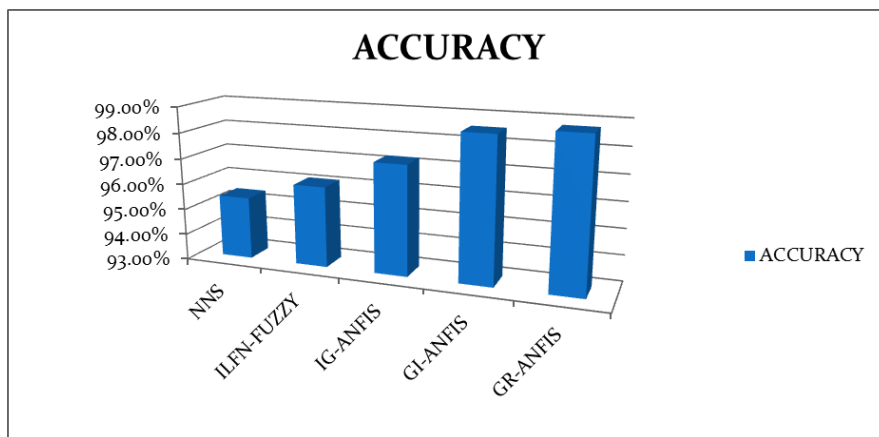
C:\061720\weka\weka
Number of Leaves : 17
Size of the tree : 32
Time taken to build model: 0 seconds
--- Evaluation on training set ---
Time taken to test model on training data: 0.01 seconds
--- Summary ---
Correctly Classified Instances      253      95.42 %
Incorrectly Classified Instances     15
Kappa statistic                     0.9701
Mean absolute error                  0.0284
Root mean squared error              0.0938
Relative absolute error              0.0476 %
Root relative squared error          23.5772 %
Total Number of Instances          268
--- Detailed Accuracy By Class ---
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| TP Rate | FP Rate | Precision | Recall | F-Measure | MDC | ROC Area | ROC Area | ROC Area | Class |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 1.000   | 0.000   | 1.000     | 1.000   | 1.000     | 1.000 | 1.000   | 1.000   | 1.000   | Junk Food |
| 0.985   | 0.021   | 0.944     | 0.985   | 0.964     | 0.951 | 0.991   | 0.994   | 0.994   | Highcholesterol Food |
| 0.981   | 0.000   | 1.000     | 0.981   | 0.990     | 0.984 | 0.997   | 0.994   | 0.994   | Oil Food |
| 0.976   | 0.004   | 0.988     | 0.976   | 0.982     | 0.973 | 0.988   | 0.988   | 0.988   | Junk Food |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
Weighted Avg. : 0.981 0.007 0.981 0.981 0.981 0.972 0.988 0.988 0.988
--- Confusion Matrix ---
+-----+-----+-----+-----+-----+-----+
| a | b | c | d | e | -- classified as |
+-----+-----+-----+-----+-----+-----+
| 2 | 0 | 0 | 0 | 1 | a = Junk Food |
| 0 | 47 | 0 | 2 | 1 | b = Highcholesterol Food |
| 0 | 2 | 184 | 0 | 1 | c = Oil Food |
| 0 | 2 | 0 | 80 | 1 | d = Junk Food |
+-----+-----+-----+-----+-----+-----+

```

The weka tool using to analysis for junk food, high cholesterol food, oil food with predict heart attack.

5. Comparison for result accuracy

THE APPROCH	ACCURACY
NNS	95.42%
ILFN-FUZZY	96.13%
IG-ANFIS	97.24%
GI-ANFIS	98.56%
GR-ANFIS	98.62%



Comparison of classification accuracy between GI-ANFIS and some previous work

Conclusion

The research used GR-ANFIS data mining technique on UCI heart data sets to provide the diagnosis results. The result from the move toward is positive. If further attempts are engaged in the application of Information Technology in diagnosing various diseases such as heart then efficient, timely and decent healthcare services will be realized. Great databases that used in the medical segment still have a concern of Missing features values brought about by many factors as discussed early. GR-ANFIS and approach had significantly expert result.

References

1. Kamber M. Data Mining: Concepts and Techniques University of Illinois at Urbana Champaign, 2006.
2. Baskar SS. Filter based feature selection approach to enhance the classification accuracy, 2013.
3. https://en.wikipedia.org/wiki/Information_gain_ratio.
4. Kaur G, Chhabra A. Improved J48 Classification Algorithm for the Prediction of Diabetes. International Journal of Computer Applications 2014;

- 98(22): 0975–8887.
6. Saeys Y, Abeel T, Peer YV. Robust feature selection using ensemble feature selection techniques, W. Daelemans et al. (Eds.): ECML PKDD 2008; 5212: 313–325.
7. Yu L, Liu H. Feature Selection for High-Dimensional Data: A Fast Correlation Based Filter Solution. Proceedings of the Twentieth International Conference on Machine Learning, ICML-03, Washington, D.C., 2003; 856-863.
8. Altidor W, Khoshgoftaar TM, Van Hulse J, Napolitano A. Ensemble feature ranking methods for data intensive computing applications. Handbook of data intensive computing, Springer Science + Business media, LLC 2011; 349-376.
9. Yen GG, Meesad P. Constructing a fuzzy rule-based system using the ILFN network and Genetic Algorithm 2001.
10. Kiage BN. A Data Mining Approach for forecasting Cancer Threats International Journal of Engineering Research and Management 2014; 1:6.