ISSN:0975-9646

Mona Shrivastava et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1688-1691

Knowledge Extraction Mechanism using Vertically Partition id3 and Fuzzy based System

Mona Shrivastava
Department of Computer Science
RKDF IST
Bhopal, India

Prof. Nirest Sharma
Department of Computer Science
RKDF IST
Bhopal, India

Prof. Gaurav Shrivastava
Department of Computer Science
RKDF IST
Bhopal, India

***Abstract*— Knowledge extraction is a technique of analyzing the dataset so that some meaningful information can be gathered from the dataset which can be used in various applications. Here in this paper an efficient technique is implemented for the knowledge extraction using combinatorial method horizontal partition decision tree and fuzzy rules based system with genetic algorithm. The proposed technique implemented provides normalized results and provides more accuracy.**

***Index Terms*—**Genetic Algorithm, Fuzzy logics, MOEA, membership function, fitness function.

I. Introduction

The make use of of genetic algorithms (GAs) to design fuzzy systems constitutes one of the branches of the Soft Computing paradigm: genetic fuzzy systems (GFSs). The most known approach is that of genetic fuzzy rule-based systems, where some components of a fuzzy rule-based system (FRBS) are derived (adapted or learnt) using a GA. A number of additional approaches within reach of include genetic fuzzy neural networks and genetic fuzzy clustering, surrounded by others.

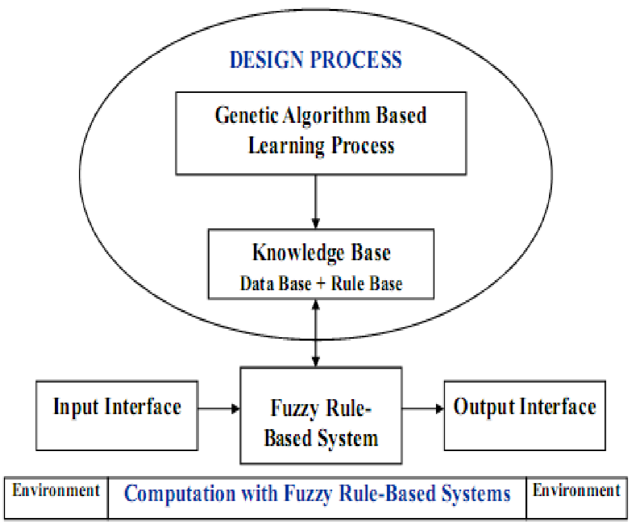

Figure: Genetic Fuzzy Rule-Based Systems

Evolutionary algorithms were not specifically designed as machine learning systems, similar to further advancement like neural networks. However, it is well known that a learning task can be modeled as an optimization problem, and thus solved through evolution. Their authoritative investigate in composite, ill-defined difficulty spaces have allowed are appropriating evolutionary algorithms effectively to an enormous variety of machine learning and knowledge discovery jobs. Their elasticity and competence to integrate obtainable knowledge are also very remarkable distinctiveness for the problem solving.

Fuzzy association rule mining discovers patterns in operations, such as shopping storage bins in a superstore, or Web page right of entries by a visitor to a Web site. Sequential patterns can be nearby in fuzzy association rules for the reason that the fundamental process producing the data can be go-ahead. On the other hand, obtainable solutions may not discover all attractive patterns for the reason that of until that time unrecognized difficulty that is making known. The appropriate meaning of fuzzy association rules alters because of the self-motivated characteristic of data. The stationary fuzzy representation and conventional search method are insufficient.

The fuzzy rule-based systems (FRBCs) are promising due to they relate to the logic of human beings and the rules they produced are easy for human to understand. To generate an understandable fuzzy systems, the concept of interpretability is taken into account to preserve semantic meaning and complexity of FRBCs. For decades, most of researchers devote to find a good trade-off between interpretability and accuracy associated with fuzzy model [1],[2]. With regard to these two criteria, multi-objective evolutionary algorithms (MOEAs) are utilized as effective global search approaches to optimize these objectives simultaneously ([3], [5]). However, by means of MOEAs, a set of non-dominated solutions with respect to different objectives, instead of a single fuzzy rule classifier are obtained. In this case, the problem of selecting the suitable and reliable solutions as final classifier has given rise to a growing interest by many researchers.

The factors defined in the first phase are explained in specify in the literature analysis; the membership functions describe how words, or linguistic terms, are make use of to express quantities. For example, Figure 1.1 shows the membership functions that describe three linguistic terms low, medium and high.

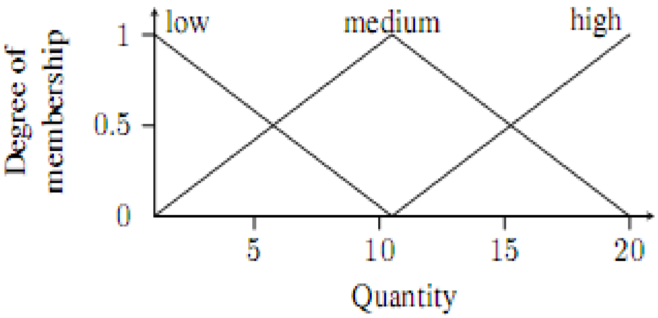

Figure: Example of membership functions.

Once these are defined, the parameters are then used for mining fuzzy association rules. On the other hand, mining fuzzy association rules that take place more commonly transversely a predetermined period of time currently identify a new problem. The conventional move toward for mining temporal fuzzy association rules take for granted the constraints are static. The membership functions, which decide the linguistic terms, do not transform all the way through the duration of the entire dataset. The appropriate meaning of linguistic terms may alteration with factors such as seasonal weather, sports events [6], or unforeseen events, e.g., hurricanes [4].

II. Related Studies To Multiobjective Genetic Fuzzy Systems

In this part, we explain a wide range of related review studies on to multi objective genetic fuzzy systems in various research areas.

Multi objective Genetic Fuzzy Rule Selection: Two-objective genetic fuzzy rule selection for the design of fuzzy classifiers [8], which is a multi objective version of weighted sum-based rule selection [7], is one of the earliest studies on multi objective genetic fuzzy systems. In [8], first a large number of candidate fuzzy rules were extracted from numerical data by a heuristic rule extraction procedure. Then an EMO algorithm was used to search for a number of non-dominated subsets of the candidate fuzzy rules with respect to accuracy maximization and complexity minimization. Let N be the number of the extracted candidate rules. A subset of the N candidate rules is represented by a binary string of length N and handled as an individual in two-objective genetic fuzzy rule selection. Since binary strings are used as individuals, we can directly apply existing EMO algorithms such as NSGA-II, SPEA and SPEA2 with standard genetic operators to two-objective genetic fuzzy rule selection. The two-objective formulation in [8] was extended to a three-objective one in [9] by giving as the number of antecedents.

Multi objective Fuzzy Genetics-Based Machine Learning: Studies on genetics-based machine learning algorithms are often divided into two classes: Pittsburgh and Michigan approaches A rule set is handled as an individual in the Pittsburgh approach while a single rule is handled as an individual in the Michigan approach. The finally obtained rule set is usually the best individual in the final population in the Pittsburgh approach while it is the final population in the Michigan approach. Another category of GBML is an iterative rule learning approach [10], where a single rule is obtained from its single execution. Thus multiple runs are required to generate a rule set in the iterative rule learning approach. Multi objective GBML algorithms have usually been implemented in the framework of the Pittsburgh approach. In general, the antecedent part of each rule is coded as a substring in Pittsburgh-style GBML algorithms. A rule set is represented by a concatenated string. A substring of integers and/or real numbers is often used to represent a single rule.

III. Literature Review

In this paper, [11] here they proposed a hybrid structure for an automatic and interpretable knowledge extraction in databases. To accomplish objective of this work is to show a system able to automatically generate Fuzzy Rule (FR) sets of computational intelligence techniques was applied in order to meet the requirements for automatic rules extraction with less human participation in order to get better classification of instances or transactions in an efficient way in direction of both numerical and interpretable results as, in some cases, it's better to have results with less efficiency in a numerical terms but with more efficiency in terms of interpretability.

In this circumstance, an interpretable consequence means a effect where the professional can know the reasons about the given arrangement by the system. To achieve the objectives above, a set of well know Computational Intelligence techniques aligned with the presented problems and objectives has been used such as Decision Trees (DT), Fuzzy Logic (FL) and Genetic Algorithms (GA). A hybrid system is proposed [11] with the capability to explore the best distinctiveness of each method in a complementary form. Some of this best distinctiveness is learning, interpretability, search and optimization. The outcome of the proposed methodology is a Genetic Fuzzy Rule Based System (GFRBS).

A GFRBS assisted by a Decision Tree was presented by the authors in this paper [12]. The following three phases were used to generate a GFRBS:

1. Generation of the Decision Tree;
2. Generation of the Fuzzy Rule-Based System; and
3. Tuning of the fuzzy membership functions using a Genetic Algorithm.

In this work, the GFRBS assisted by a Decision Tree is adapted to accommodate the PHM problem. The first challenge of the proposed solution is to generate an accurate RUL prediction model. The second challenge is to generate a fuzzy model with low complexity such that its interpretability is improved.

IV. Proposed Methodology

Generation of decision tree using vertical partition based id3 decision tree

Input Layer:

1. First of all the whole dataset is partitioned vertically into number of parties.
2. Let S1, S2, S3, S4……Sn be the ‘N’ no. of parties from the Dataset ‘D’.

A1 attribute	A2 attribute	A3 attribute	A4 attribute
Valu1	Valu2	Valu3	Valu4
Valu5	Valu6	Valu7	Valu8
Valu9	Valu10	Valu11	Valu12
Valu13	Valu14	Valu15	Valu16
Valu17	Valu18	Valu19	Valu20
Valu21	Valu22	Valu23	Valu24

Original Full Dataset

A1 attribute	A2 attribute
Vaul1	Valu2
Valu5	Valu6
Valu9	Valu10
Valu13	Valu14
Valu17	Valu18
Valu21	Valu22

Party 1

A3 attribute	A4 attribute
Valu3	Valu4
Valu7	Valu8
Valu11	Valu12
Valu15	Valu16
Valu19	Valu20
Valu23	Valu24

Party 2

3. Now calculate Information gain for each of the parties using

$$I(y, n) = -\frac{T(y)}{T(y) + T(n)} - \frac{T(n)}{T(y) + T(n)}$$

4. After calculating Information for each of the parties, now calculate Entropy for each of the attribute for each of the parties.
5. Now calculate Gain for each of the attribute for each of the parties and send them to the certified authority.
6. The certified authority then adds the gain from the two parties and checks the highest gain for the attribute.
7. Create decision tree with best gain attribute at the root node and repeat the steps for the remaining attributes.

Generation of fuzzy rule based system

The process is simplified because the variables in the first phase were previously categorized. The fuzzy rules can be extracted from the decision tree rules by travelling each path of the tree from the leaf to root node. In this phase the algorithm is concerned only in the fuzzy rule extraction.

Genetic Algorithm

Begin
1. t=0
2. Initialize population of the genetic algorithm as pop(t).
3. Now calculate the fitness value of the dataset.
4. next cycle
5. if child is less than threshold value
 move to step 10
6. select the last two child’s.
7. Apply crossover p(t)
8. Now mutate p(t)
9. go to step 3
10. output the child’s with maximum fitness and stop
End

V. Result Analysis

The proposed technique implemented here is applied on appendicitis dataset which contains 7 attributes and two classes 0 and 1. The dataset is used to generate a number of rules using hybrid combination of decision tree and fuzzy system and genetic algorithm. The results shown below are the analysis of computational time when applied on different decision tree.

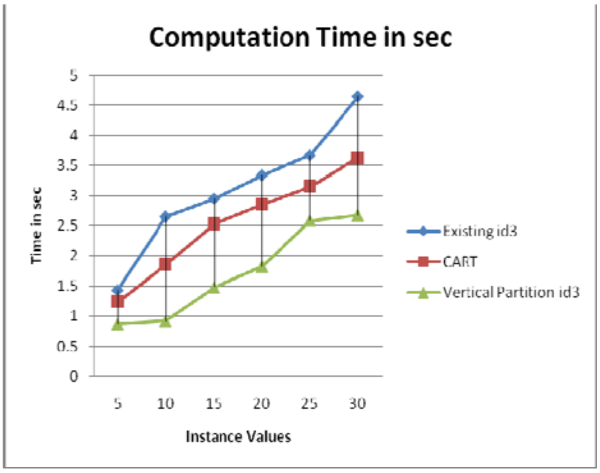

Figure 1. Result Analysis of the computational time

The figure shown below is the decision tree created when using dataset containing 7 attributes. The decision tree shows the dependency of the attributes through which the fuzzy rules are generated.

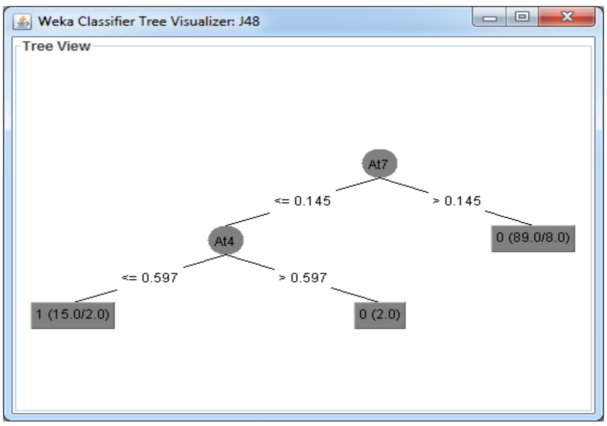

Figure 2. Decision tree created from proposed work

The table shown below are the number of rules generated when applied proposed algorithm on the appendicitis dataset. There are total 3 rules generated from the proposed techniuqe.

Rules Generated from Proposed Algorithm
If (At7 is high) then (class is 0)
If (At7 is low) and (At4 is high) then (class is 0)
If (At7 is low) and (At4 is low) then (class is 1)

Table 1. Generation of rules from proposed work

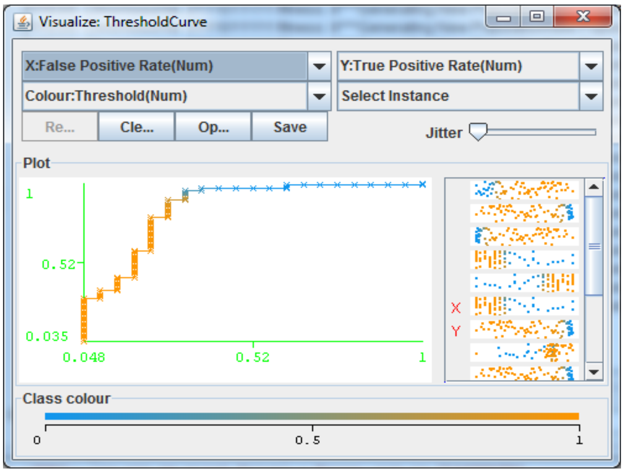

Figure 3. Analysis of False Positive Rate

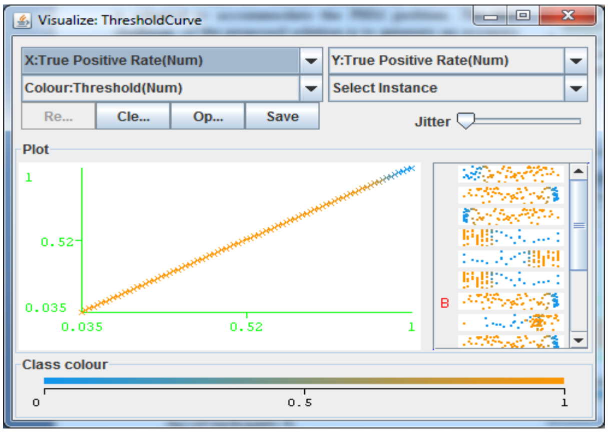

Figure 4. Analysis of True Positive Rate

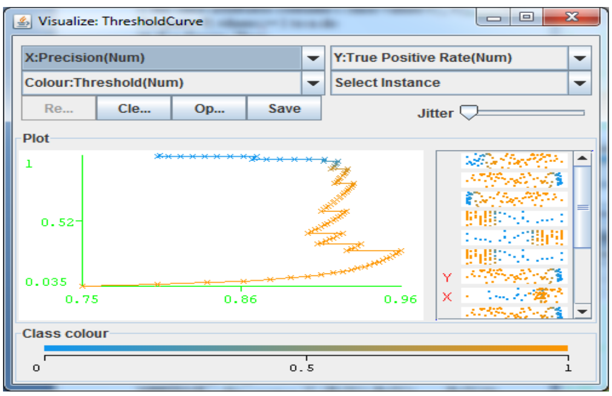

Figure 5. Analysis of Precision

VI. CONCLUSION

The proposed methodology implemented here is efficient in terms of rules generation from decision tree and provides more accuracy and less error rate as compared to the other existing techniques of knowledge extraction. The proposed methodology is tested on appendicitis dataset and after applying methodology on the above dataset it performs better and gives efficient results.

REFERENCES

[1] Corrado Mencar and Anna Maria Fanelli. Interpretability constraints for fuzzy information granulation. Inf Sci., 178(24):4585-4618, 2008.

[2] Shang-Ming Zhou and John Q. Gan. Low-level interpretability and highlevel interpretability: a unified view of data-driven interpretable fuzzy system modelling. Fuzzy Sets and Systems, 159(23):3091-3131, 2008.

[3] MJ. Gacto, R. Alcala, and F. Herrera. Integration of an index to preserve the semantic interpretability in the multi objective evolutionary rule selection and tuning of linguistic fuzzy systems. Fuzzy Systems, IEEE Transactions on, 18(3):515-531, 2010.

[4] Leonard, D. (2005) After Katrina: Crisis Management, the Only Lifeline Was the Wal-Mart. FORTUNE Magazine, (October 3).

[5] R. Y Zhang, X. Wu, Z. Xing and WL Hu. On generating interpretable and precise fuzzy systems based on pareto multi-objective cooperative co-evolutionary algorithm. Applied Sof Computing, 11(1):1284-1294, 2011.

[6] Saleh, B. and Masseglia, F. (2010) Discovering frequent behaviors: time is an essential element of the context. Knowledge and Information Systems, 28, pp. 1–21.

[7] M. H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, “Selecting fuzzy if-then rules for classification problems using genetic algorithms,” IEEE Trans. on Fuzzy Systems 3 (3) 260-270, 1995.

[8] H. Ishibuchi, T. Murata, and I. B. Turksen, “Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems,” Fuzzy Sets and Systems 89 (2) 135-150, 1997.

[9] H. Ishibuchi, T. Nakashima, and T. Murata, “Three-objective genetics-based machine learning for linguistic rule extraction,” Information Sciences 136 (1-4) 109-133, 2001.

[10] L. Castillo, A. Gonzalez, and R. Perez, “Including a simplicity criterion in the selection of the best rule in a genetic fuzzy learning algorithm,” Fuzzy Sets and Systems 120 (2) 309-321, 2001.

[11] Rogério Ishibashi, Cairo Lúcio Nascimento Júnior, “Knowledge Extraction using a Genetic Fuzzy Rule-based System with Increased Interpretability” SAMI 2012 • 10th IEEE Jubilee International Symposium on Applied Machine Intelligence and Informatics • January 26-28, 2012.

[12] Ishibashi, R.; Nascimento Jr., C. L.; "Knowledge extraction using a genetic fuzzy rule-based system with increased interpretability". 2012 IEEE 10th International Symposium on Applied Machine Intelligence and Informatics (SAMI 2012), pp. 247-252, 26-28 January 2012.