International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

Performance of Fuzzy Rough Sets and Fuzzy Evolutionary Classifiers using Medical Databases

S. Poongothai, C. Dharuman, P. Venkatesan

***Abstract:** As the technology improving, the problems of mankind, regarding health issues also increasing day by day. Nowadays high dimensionality data are available for various health problems which is very difficult to handle manually. The aim of this paper is to construct algorithms for extracting the relevant information from the large amount of data and classifying using various hybrid techniques like Fuzzy-Rough set and Fuzzy Evolutionary Algorithms. The efficiency of Fuzzy classifiers has been improved by hybridization method. This paper proposes a comparison of fuzzy hybrid techniques like Fuzzy Rough set and Fuzzy EA for the diagnosis of Hepatitis taken from UCI repository. The results of comparison and classification shows that the proposed technique performs better than other normal methods.*

***Keywords:** Fuzzy logic, Rough sets, Evolutionary algorithms, Hybrid techniques*

I.INTRODUCTION

In machine learning area, one of the major problems is classification. It is based on supervised learning process. From this process, rules can be framed which is used for prediction. When the number of attributes are raises, the number of rules also get increased [1].The classification rate suffers when the number of training data cases is smaller than the number of attributes [2]. In this paper, we focus on hybridization of fuzzy with rough set and genetic algorithms and classification done.

II.MACHINE LEARNING METHODS

Fuzzy Sets and Logic

The idea of Fuzzy set was developed by Mathematician Lofti Zadeh in 1960's, is a best tool in mathematics to manage the uncertainty that arises due to impreciseness for solving the problem. Mathematically, a fuzzy set A is defined to be a set of ordered pairs and is written as

$A = \{(x, \mu_A(x)) \mid x \in X\}$,

Where X is the universe of discourse and $\mu_A(x)$ is called the membership function of x in A. In order to deal with reasoning problem, multi valued logic that is derived from Fuzzy set theory known as Fuzzy logic has been applied [3].

Manuscript published on 30 August 2019.
***Correspondence Author(s)**
S. Poongothai, Asst. Professor, Department of Mathematics, SRM University, Ramapuram Campus, Chennai, India
C. Dharuman, Professor, Department of Mathematics, SRM University, Ramapuram Campus, Chennai, India
P. Venkatesan, Faculty of Research, Sri Ramachandra University, Chennai, India

The main reasons for the development of fuzzy systems are [4]: (i) Human reasoning is very simple to adapt using fuzzy system; (ii) Mathematically, models of societal problems can easily be constructed; (iii) creates transition between members and non-members easily; (iv) system fluctuations are less sensitivity; and (v) Description or linguistic rules can be easily framed by fuzzy system. In this paper, FURIA (Fuzzy Unordered Rule Induction Algorithm) is considered for classification. Based on RIPPER algorithm, FURIA is developed. This method is well versed in differentiating and separating the classes and so there is no use of default rule and the classes order are irrelevant [5].

Rough Sets

In 1980's Zdzislaw Pawlak [6] proposed a new method called Rough set which is another mathematical tool used to handle vagueness and imprecise. In this approach, uncertainty and impreciseness are not expressed by the membership functions as in fuzzy sets but defined by the boundary region of a set. The major difference of these two approaches are fuzzy sets are employed with membership function, where Rough sets are defined by topological operations called upper approximations and lower approximations, that requires advanced mathematical concepts. It is used to find the reduct set of attributes from large set of data present in the decision system. This reduct set then used to frame rules for classification through data mining techniques. Let U be the universe, a nonempty finite set of M objects, $\{y_1, y_2, y_3, \ldots y_m\}$, Q be the finite set of attributes, then the information system is defined as

$$IS = \langle U, Q, V, f\rangle$$

where $V = \bigcup_{q\in Q} V_q$ i.e., V_q is the domain for attribute q and $f: U \times Q \to V$ is the information function such that $f(y,q) \in V_q$ for every $q \in Q, y \in U$.

Indiscernibility Relation (IR) on U is defined as

$$IR(A) = \{(x,y) \in U: for\ all\ a \in A, f(x,a) = f(y,a)\}$$

Where $A \subseteq Q$ is the set of attributes. The decision table is defined as $DT = \langle U, C \cup D, V, f\rangle$, where C and D are the condition attribute and the decision attribute respectively, $V = \bigcup_{q\in C\cup D} V_q$ and $f: U \times (C \cup D) \to V$ is the decision function such that $f(y,q) \in V_q$ for every $q \in Q, y \in U$. For a given space S, a subset $A \subseteq Q$ determines the approximation space as $AS = (U, IR(A))$ in S. For a given $A \subseteq Q$ and $Y \subseteq U$, the upper and lower approximations of AS are defined respectively as

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

$\overline{A}Y = \{y \in U: [y]_A \subseteq Y\} = \cup\{X \in A^*: X \subseteq Y\}$ and

$\underline{A}Y = \{y \in U: [y]_A \cap Y \neq \phi\} = \cup\{X \in A^*: X \cap Y \neq \phi\}$

The two major concepts of Rough set theory are core and reduct which is used for selecting feature attributes as well as reducing the attributes. In the information system, some attributes are redundant with respect to A^* that is generated by attributes $A \subseteq Q$. Based on attributes qualities, it is possible to find irrelevant attributes and also we can remove those attributes, without affecting the classification accuracy. If $IR(A) = IR(A - \{a\})$, then the attribute $a \in A \subseteq Q$ is considered as dispensable in the set A i.e., indiscernibility relation of A and $A - \{a\}$ are same. If it is not identical, then a is indispensable in A. These dispensable attributes are not used to improve the accuracy of classification of IS. All indispensable attributes together forms the set called Core of A and it is denoted by $\mathfrak{C}(A)$.

If all the attributes in the set $A \subseteq Q$ are indispensable, then the set A is called *orthogonal*. Also if there exists a proper subset $E \subset A$ that is orthogonal and also preserves the classification, then it forms the reduct set generated by A. Hence a reduct set of A, is denoted by $\mathfrak{R}(A)$, is defined as $\mathfrak{R}(A) = (E \subset A, IR(A) = IR(E), E \text{ } is \text{ } orthogonal)$, where E is a reduct of A [6]. Also the reduct set cannot be further reduced.

Evolutionary Algorithms

In 1960 I. Rechenberg introduced the idea of Evolutionary Algorithms in his work "*Evolutionary strategies*". EAs are algorithms of general class stochastic optimization based on neo-Darwinian theory. The main idea behind EAs are *Survival of the Fittest* [7]. Most common Evolutionary Operators are Selection, Crossover and Mutation. The main four streams of EAs are (i) Evolutionary Programming by Fogel et al., in 1966 (ii) Evolutionary Strategies by Rechenberg in 1973 (iii) Genetic Programming by Holland in 1975 (iv) Genetic Algorithms by Holland in 1975. In the proposed method genetic algorithms is used for feature selection. It is a global search technique applied widely to optimize the problems (Holland, 1975) [8]. It can also be used for solving the problems having objective function as discontinuous, stochastic, non linear or non-differentiable [9].

III.ALGORITHMS FOR HYBRIDIZATION

Fuzzy Rough Sets

Fuzzy Rough sets is the generalization of rough set s having fuzzy background i.e., to obtain results of fuzzy in a crisp space [10-12]. This model is mostly utilized in knowledge acquisition with decision tables of real valued conditional attributes and reasoning [13–20]. In this case the decision attribute values are termed as fuzzy and the conditional attribute values are termed as crisp value. The major role of this model is to define the lower and upper approximation of the set. Let U and V be two nonempty universe discourse set and R be a fuzzy relation from U to V, then the generalized fuzzy approximation space is denoted by the triple (U, V, R). When U equals V and R defined a fuzzy relation on U, then (U, R) is termed as a fuzzy approximation space. The membership value of 1, 0.5 and 0 are for elements having lower approximation (positive region), boundary region and upper approximation (negative region) respectively [21]. Using Fuzzy partition, the input set X is partitioned into N clusters named $C1, C2, \ldots, CN$ be generated [22]. Each equivalence class having output classes in various forms is denoted by each cluster. These output classes are identified by fuzzy upper and lower approximations equivalence classes. Let (U, R) be a approximation space and μ ∈ F(U). The lower is expressed as $\underline{R}$ (μ) and upper rough approximations of μ in (U, R) are expressed as $\overline{R}$ (μ), which are fuzzy subsets in U defined by

$$\underline{R}(\mu)(x) = \wedge\{\mu(y);\ y \in [x]_R\},$$

$$\overline{R}(\mu)(x) = \vee\{\mu(y);\ y \in [x]_R\},$$

for all x ∈ U. μ is called definable in (U, R) if $\underline{R}$ (μ) = $\overline{R}$ (μ), otherwise it is a rough fuzzy set [23]. Fuzzy-Rough Feature Selection (FRFS) used effectively for real valued or discrete noisy data to reduce the attributes without the need for user-supplied details. Each feature required to be in the form of fuzzy partition that can be derived from the data itself.

Fuzzy Genetic Algorithm

The power of Fuzzy Rule Based Systems (FRBS) which is an extension of classical rule based systems that are dealing with "IF-THEN" rules, its antecedents and consequents are represented by fuzzy logic statements are described for solving modelling problems, control problems, data mining problems [24-26], in numerous application areas. Fuzzy genetic algorithm is a fuzzy system augmented by the process of learning based on evolutionary computation. A *fuzzy representation* is proposed for handling the optimization problems of parameters whose variables are continuous domains. Each parameter in the problem is associated with a number (m) *fuzzy decision variables* which belongs to the interval [0,1] [27]. Each parameter is linked with the values of the decision variables to get the solutions. The workflow of this process is shown in figure 1.

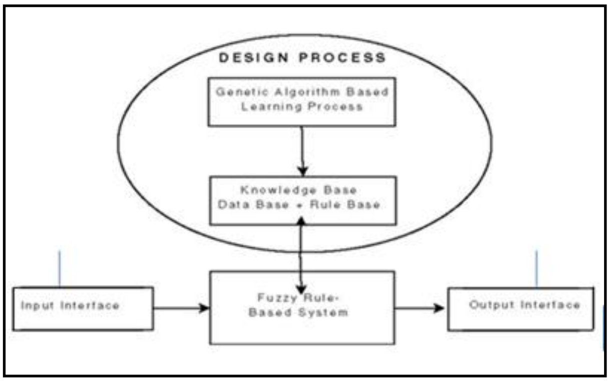

Fig. 1 Fuzzy Genetic method

Retrieval Number: J10660881019/19©BEIESP
DOI: 10.35940/ijitee.J1066.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-10, August 2019

Description of datasets

The proposed method is used to classify the dataset of Hepatitis disease taken from UCI machine learning repository. The number of instances and attributes are 155 and 20 respectively. The aim is to classify the death and live due to presence of hepatitis [28].

IV.RESULTS

In the proposed method, each attribute Ai in a rule can be expressed in the form ($A_i \in I$), if the selector is $Ai \le v$, then I is an interval (-∞, v], if the rule selector is $Ai \ge u$ then I is an interval [u,∞), and I = [u, v] if it contains both $Ai \le v$ and $Ai \ge u$, the fuzzy intervals obtained by trapezoidal membership function replaces the intervals [5]. For rough set, the data which has taken is stored in a table, called decision table. Rows represents objects and columns represents attributes. Rough set theory defines three regions namely lower, upper and boundary approximations. All the objects that are positively classified contained in Lower approximation, the probably classified objects contained in upper approximation, while the boundary is the difference between these two approximations. But in the method of GA, by selecting individuals (attributes) from the current population (Dataset) randomly and uses them to produce new offsprings for the next generation. After applying crossover and mutation, several generations obtained by GA, it tries to give the optimal solution. At last the lowest fit individuals in the original population are replaced by the newly created offsprings. This replacement always gives the best set of individuals deleting the worst ones. The solution obtained by each generation is better than its previous one. The iterations can be continued till the stopping criteria or the desired result obtained [5].In this paper, Fuzzy, Fuzzy rough set (FRS) and Fuzzy Genetic Algorithms (FGA) are used to classify the database. Using the FRS the attributes are reduced to 13 and it is shown in figure 2. The classification rate for this system is 81%.

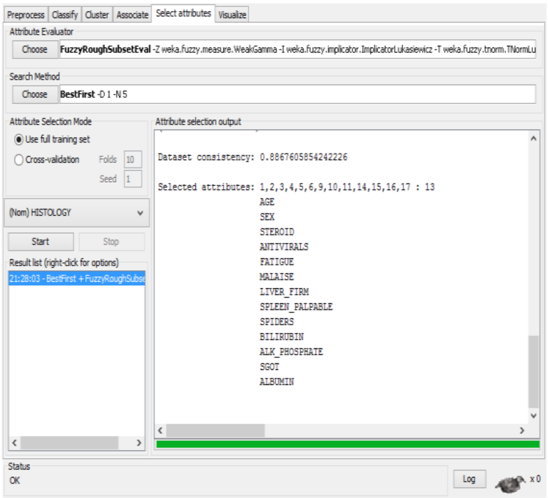

Fig. 2 Fuzzy rough feature selection

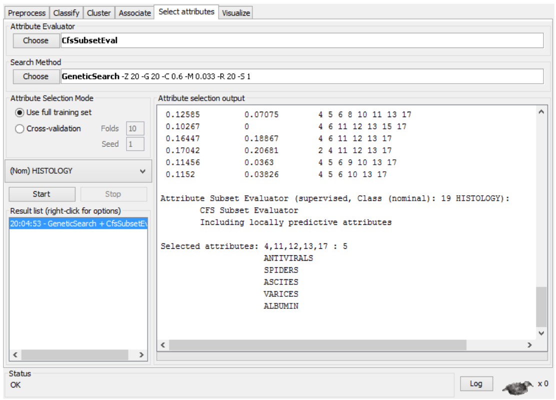

Fig. 3 Feature selection by GA

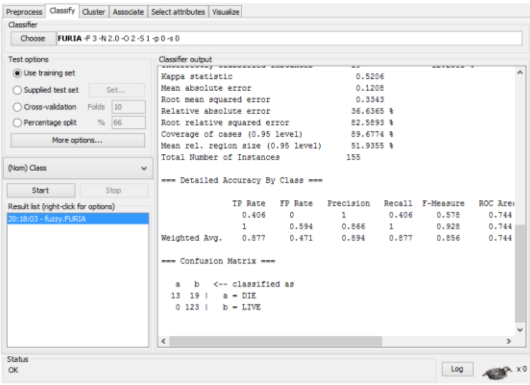

Fig. 4 Confusion matrix of fuzzy classification

By the second method the attributes are reduced to 5 and 87% correctly classified. Reduced attributes by GA is shown in figure 3. The confusion matrix of this model is given in figure 4. Also the classification of Fuzzy alone has been discussed and given in figure 5 which gives only 79.35% classification.

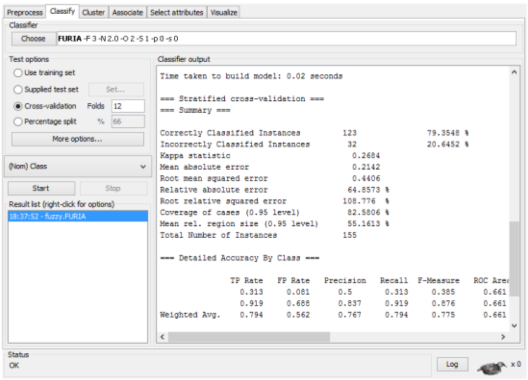

Fig. 5 Classification of Fuzzy

V.DISCUSSIONS AND CONCLUSIONS

In this paper, fuzzy rough and fuzzy genetic methods are compared for classification for Hepatitis. Also along with these hybrid methods, Fuzzy is compared and this paper has shown that the idea of hybridization of genetic algorithm with fuzzy sets in a fruitful way. The following table 1 shows the comparison of the proposed model of the paper.

Table. 1 Comparison of Classification

Method of Classification	No. of Attributes	Classification rate (%)
Fuzzy	20	79.35
FRS	13	81
FGA	5	87

This shows hybridization role plays in a better way compared to concept of fuzzy alone. Out of these two hybrid methods, FGA is good in classification. The idea of differential evolutionary algorithm shall be combined with fuzzy sets for future works.

REFERENCES

1. R. Bellman, Adaptive Control Processes: A Guided Tour, Princeton University Press,961.
2. G.V. Trunk, .A Problem of Dimensionality: A Simple Example,. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(3), 306-307, 1979.
3. Zadeh L .A, Fuzzy sets, Information and Control, Vol 8, 1965, pp. 338–353.
4. K.B.Mankad, and P.S.Sajja, Measuring human intelligence by applying soft computing techniques: A genetic fuzzy approach , in Diagnostic Test Approaches to Machine Learning and Common-sense Reasoning Systems, X.A. Naidenova and D. Ignatov, Eds.
5. Hershey, PA: IGI Global, pp.128-153,July 2012. [Online].Available:http://www.irmainternational.org/viewtitle/69407/ [Accessed Dec 2014].
6. del Jesus, M., Hoffmann, F., Navascues, L., and Sánchez, L. (2004). Induction of fuzzy rule- based classifiers with evolutionary boosting algorithms. IEEE Transactions on Fuzzy Systems, **12**(3), 296.308.
7. Pawlak Z. (1982): Rough sets. | Int. J. Comp. Sci., Vol.11, pp.341-356. Rechenberg ,"Evolutionary strategies" 1960.
8. Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. Ann Arbor: University of Michigan Press. (A 2nd edn was published in 1992. MIT Press, Cambridge 183 pp.
9. Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley 412 pp.
10. D.Boixader, J.Jacas, J.Recasens, Upper and lower approximations of fuzzy sets, International Journal of General Systems 29 (2000) 555–568.
11. D.G.Chen, S.Y.Zhao, Local reduction of decision system with fuzzy rough sets, Fuzzy Sets and Systems 161(2010)1871–1883.
12. M.D.Cock, C.Cornelis, E.E.Kerre, Fuzzy rough sets: The forgotten step, IEEE Transactions on Fuzzy Systems15(2007)121–130.
13. C.Cornelis, G.Deschrijver, E.E.Kerre, Implication inintuitionistic fuzzy and interval- Valued fuzzy set theory: Construction, classification, application, International Journal of Approximate Reasoning35(2004)55–95.
14. Q.H.Hu, D.R.Yu, W.Pedrycz, D.G.Chen, Kernelized fuzzy rough sets and their applications, IEEE Transactions on Knowledge and Data Engineering 23(2011)1649– 1667.
15. Q.H.Hu, L.Zhang,S. An, D.Zhang, D.R.Yu, On robust fuzzy rough set models, IEEE Transactions on Fuzzy Systems 20 (2012) 636–651.
16. Q.H.Hu, L.Zhang, D.G.Chen, W.Pedrycz, D.R.Yu, Gaussian kernel based fuzzy rough sets: Model, uncertainty measures and applications, International Journal of Approximate Reasoning 51 (2010) 453–471.
17. R.Jensen, Q.Shen, Fuzzy-rough attributes reduction with application to web categorization, Fuzzy Sets and Systems 141(2004)469–485.
18. R.Jensen, Q.Shen, Semantics-preserving dimensionality reduction: Rough and fuzzy rough based approaches, IEEE Transactions on Knowledge and Data Engineering 16 (2004) 1457–1471.
19. R.Jensen, Q.Shen,Fuzzy-rough sets assisted attribute selection, IEEE Transactions on Fuzzy Systems 15(2007)73–89.
20. R.Jensen, Q.Shen, New approaches to fuzzy- rough feature selection, IEEE Transactions on Fuzzy Systems 17 (2009) 824–838.
21. Kasemsiri, W., & Kimpan, C. (2001). Printed thai character recognition using fuzzy- rough sets. In TENCON 2001. Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology (Vol. 1, pp. 326-330). IEEE..
22. Rajen B. Bhatt, M.Gopal. On Fuzzy Rough Set Approach To Feature selection, IEEE Transaction on Fuzzy Systems, volume 15, pages 73-89, 2007.
23. D. Dubois, H. Prade, Rough fuzzy set and fuzzy rough sets, International Journal of General Systems 17 (1990) 191–209.
24. Palm, R., Driankov, D.: Model based fuzzy control. Springer, Berlin (1997)
25. Pedrycz, W. (ed.): Fuzzy modelling: paradigms and practice. Kluwer Academic Press, Dordrecht (1996).
26. Kuncheva, L.: Fuzzy classifier design. Springer, Berlin (2000)
27. Voigt, H.M.: Fuzzy evolutionary algorithms. Technical Report tr-92-038, International Computer Science Institute (ICSI), Berkeley (199UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/index.html.

Retrieval Number: J10660881019/19©BEIESP
DOI: 10.35940/ijitee.J1066.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)