International Journal of Engineering & Technology, 7 (4.19) (2018) 104-108

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Rule based Hybrid Weighted Fuzzy Classifier for Tumor Data

D. Winston Paul[1*], S. Balakrishnan[2], A.Velusamy[3]

[1, 2, 3] Department of Information Technology,
Sri Krishna College of Engineering and Technology, Coimbatore, Tamilnadu, India
**Corresponding Author Email: [1]winstonpauld@skcet.ac.in*

Abstract

Examination of gene based information has turned out to be so essential in biomedical industry for assurance of basic ailments. A fuzzy rule based classification is a standout amongst the most mainstream approaches utilized as a part of example arrangement issues. The fuzzy rule based classifier creates an arrangement of fuzzy if-then decides that empower exact non-straight order of information designs. In spite of the fact that there are different techniques to create fluffy if-then guidelines, the advancement of lead producing process is as yet an issue. Here, we introduce a half and half weighted fluffy order framework in which few fluffy if-then principles are chosen by methods for offering weights to preparing designs. Further, we utilize a genetic algorithm (GA) to streamline the classifier for quality articulation investigation

Keywords: *Data mining, classificaton, Bioinformatics, Fuzzy sytems ,genetic algorithms, weighted rule.*

1. Introduction

In the process of DNA analysis, microarray studies have received a lot of attention, industrially as well as in research. The gene expression analysis plays an inevitable role of both identification and prediction of DNA sequences. Microarray trials can either screen every quality a few times under fluctuating conditions or examine the qualities in a solitary situation however in various sorts of tissue [1], [2].

Here, we center around the arrangement of quality articulations. "Information mining has pulled in a lot of consideration in the data business and in the public arena in general as of late, because of the wide accessibility of gigantic measures of information and the up and coming requirement for transforming such information into helpful data and learning [5]". Arrangement is an undertaking that happens much of the time in regular day to day existence. Basically it includes separating up objects with the goal that each is relegated to one of various commonly thorough and select classifications known as classes [6].

In this paper, we utilize mixture weighted fluffy manage based characterization framework for dissecting microarray articulation information. Fluffy frameworks in view of fluffy if-then principles have been connected to different control issues. One favorable position of such fluffy framework is their interpretability. As of late fluffy manage based framework have likewise been connected to design characterization issues [4]. These are numerous ways to deal with consequently create fluffy if-then guidelines from numerical information for design grouping issues, for example, Michigan, Pittsburgh or Iterative lead learning approaches. In the proposed work, we enhance the Michigan based fuzzy rule scheme by adding weights for training patterns. Thus we can reduce the number of rules generated exponentially there by reducing number of attributes involved. Further, using genetic algorithm (GA) [11] we can make "a compact classifier for gene expression analysis".

2. Background

2.1. Fuzzy Rule Based Classification

Pattern classification normally is a directed procedure where, in light of known information, anticipate choice for new information. The known information is called preparing set and the new information is called test information. Different techniques have been proposed for fluffy order. "Let us assume that our pattern classification problem is an n-dimensional problem with C classes and m given training patterns"

$Y_p = (y_{p1}, y_{p2}, \ldots, y_{pm}), p=1, 2, \ldots, m.$

"Without loss of generality, each attribute of the given training patterns is normalized into a unit interval [0, 1]; That is, the pattern space is an n-dimensional unit hypercube $[0, 1]^n$". In this study we use "fuzzy if-then rules of the following type as a base of our fuzzy rule-based classification systems:

Rule R_j: If y_1 is A_{j1} and ... and y_n is A_{jn} then
Class C_j with CF_j, $j=1, 2, \ldots, N$

where R_j is the label of the jth fuzzy if-then rule, $A_{j1}, \ldots, A_{jn}$ are antecedent fuzzy sets on the unit interval [0, 1], C_j is a consequent class (i.e. one of the given classes), CF_j is the grade of certainty of the fuzzy if-then rule R_j, and N is the total number of fuzzy if-then rules". As "antecedent fuzzy sets, we use triangular fuzzy sets as in Fig.1and Fig.2 where we show various partitions of the unit interval into a number of fuzzy sets" [3].

The "Fig.1 shows triangular membership function and Fig.2 Shows the triangular membership function based on five fuzzy set". So, it contains five classes and 25 rules are generated (if the number of attribute is 2) for each attribute.

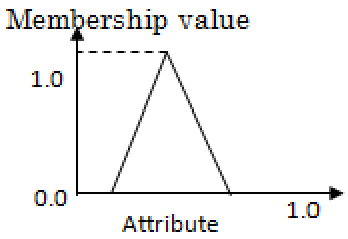

Fig. 1: Triangular membership Function

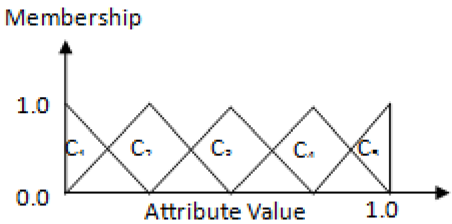

Fig. 2: Five Fuzzy Set Membership Function

2.2. Fuzzy if-then Rule Generation

Step 1: Calculate βclass h (Rj) for Class h as

$$\beta_{class\ h}(R_i) = \sum_{X_p \in Class\ h} \mu_j(y_p)/N_{Class\ h}$$

Where, μj(yp) = μj1(yp1)….. μjn(ypm).
P=1, 2, …, m. h=1,2,…,C.

Step 2: Find Class ĥ that has the maximum value of βclass h (Rj),

$$\beta_{class\ \hat{h}}(R_i) = \max_{1\le k \le C}\{\beta_{Class\ k}\ (R_j)\}$$

where βclass h (Rj) is the "average of the compatibility grades of training patterns in Class h with the fuzzy if-then rule Rj and Nclass h is the number of training patterns which their corresponding class is Class h and each of the fuzzy rules in the final classification has a certainty grade, which denotes the strength of that fuzzy rule". This number is calculated according to,

$$CF_i = \beta_{Class\ \hat{h}}(R_j) - \beta \Big/ \sum_{h=1}^{c} \beta'_{class\ h}\ (R_j)$$

Where,

$$\beta' = \sum_{h-h'} \frac{\beta_{class\ h}}{c-1}$$

2.3. Fuzzy Rule Mapping

"When a rule set S is given, an input pattern Y=(y1,y2,…,yn) is classified by a single winner rule Rj in S, which is determined as follow:

$$\mu_j\left(y_p\right).CF_j = \max\{\mu_j\left(y_p\right).CF_j \mid R_j \in S\}$$

That is the "winner rule has the maximum product of compatibility and the certainty grade CFj". The "classification is rejected if no fuzzy if-then rule is compatible with the input pattern Yp". The generation of each fuzzy rule is accepted only if its consequent class is the same as its corresponding random pattern class. Otherwise, the generated fuzzy rule is rejected.

3. Related Work

There are several approaches for classifying gene expression data. In [1], the clinical research approach method is used to categorize different types of cancer tissues and to identify new cancer classes. The approach [1] "used here is based on gene expression monitoring by DNA microarrays and are applied to human leukemia's as test case". The gene expression databases are pillars of the research [2] and mining from these databases is a major task.

In section II, we have explained the generic fuzzy classification system as expressed in [3], [7]. An "n-dimensional classification problem is analyzed using a hybrid fuzzy method". The hybrid fuzzy classification system is formed by combining the fuzzy classification scheme and genetic algorithm [10]-[12]. The "fuzzy classification scheme is used to generate the fuzzy if-then rules, which increases the number of rules exponentially with the number of attributes involved".

In order to reduce the rule generation, the modern genetic algorithm is used. Some other methods have also been applied using rule weights and weights for training patterns to reduce the rules by avoiding unimportant attributes [7], [8]. Applying weights for training patterns prove to be better than rule weights [7]. However, optimization is not achieved completely.

The main contribution of this paper is hybridizing the evolutionary genetic algorithm with weighted training patterns for optimizing the rule generation. As a result, the searching complexity of rule mapping process is reduced.

4. Hybrid Weighted Fuzzy Method

In our fuzzy rule generation approach, the weight is given for each training pattern. The initial population is generated from the weighted training patterns and is subjected to fitness evaluation. To achieve more compactness for classifier, the current population is further subjected to genetic operations.

4.1. Weights for Training Patterns

The motivation behind weight is to give need for all preparation designs. The heaviness of preparing examples can be seen as a critical factor of the examples. There are different techniques accessible for weight task, for example, class-based weighting strategy, cover based weighting technique and arbitrary weighting technique and so on. Here we underline on class based weighting technique.

The class-based weighting technique is to make an inclination toward the grouping of examples from a specific class. For instance, if the predisposition is toward the arrangement of Class1 designs, characterization frameworks are relied upon to accurately order Class 1 designs regardless of whether the quantity of misclassification/dismissal is substantial for different classes.

In this "weighting method, a weight for the pattern Yp is determined by following equation":

$$w_p = \begin{cases} 1.0 & if\ class\ of\ x_p\ is\ to\ be\ emphasized \\ 0.5 & Otherwise \end{cases}$$

4.2. Rules from Weighted Training Patterns

Let us assume that we have m training patterns Yp= (yp1,yp2, …, ypm), p=1,2, …,m. Then the rule procedure is,

[Step 1]: Calculate βclass h (Rj) for Class h as

$$\beta_{class\ h}(R_j) = \sum_{x_p \in class\ h} \mu_j\left(y_p\right).w_p/N_{class\ h}$$

h=1,2, …. C

where, $\mu j(yp) = \mu j1(yp1) \ldots\ldots \mu jn(ypm)$.

[Step 2]: Find Class ĥ that has the maximum value of βclass h (Rj),

$$\beta_{class\ \hat{h}}(R_j) = \max_{1 \le k \le C}\{\beta_{class\ k}(R_j)\}$$

The remaining procedures are same as explained earlier.

4.3. Architecture

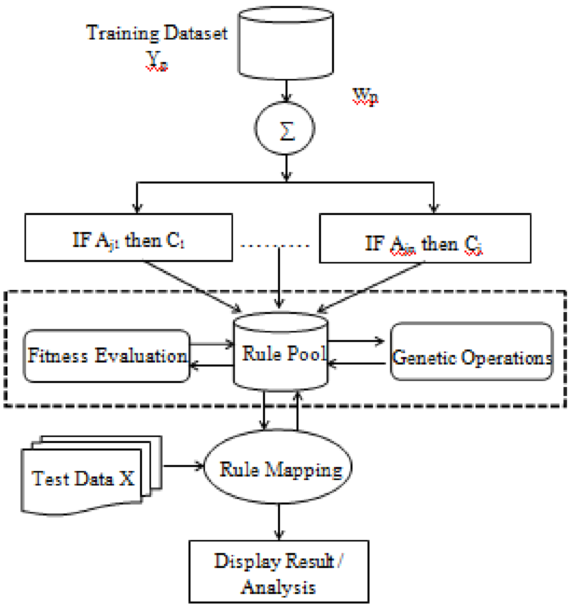

Fig. 3: Hybrid weighted fuzzy classifier

The Fig.3 shows the overall structure of the hybrid weighted fuzzy if-then rule based classification system. The known classified data is taken as input for training process. The weight is multiplied with each training pattern and the summation is given for internal input for generating rules. The rules are stored in rule pool and each rule is submitted to fitness evaluation and then for genetic operation, to remove worst rules. Finally, the unknown test data is given for classification.

4.4. Genetic Algorithm

Genetic algorithms are roused by Darwin's hypothesis of development. It is a cutting edge calculation for taking care of any enhancement issue. Answer for an issue illuminated by hereditary calculations utilizes a developmental procedure. Calculation begins with arbitrarily producing the arrangement of conceivable answers for an advancement issue called populace. Arrangements from one populace are taken and used to shape another populace. This is inspired by an expectation, that the new populace will be superior to the old one. At that point arrangements, which are chosen to frame new arrangements, are chosen by their wellness - the more appropriate they are the more shots they need to imitate. This is rehashed until the point when some condition is fulfilled. The hereditary tasks are determination, hybrid and transformation. The new guidelines are created from the principles in the present populace utilizing hereditary activities. To perform hybrid task, two fluffy if-then standards are arbitrarily chosen from the present populace and the better administer with the higher wellness esteem is picked as a parent string. A couple of parent strings is picked by repeating this strategy twice. From the chose match of parent strings, two new strings are produced by a hybrid activity. The "crossover operator is applied to each pair of parent strings with a pre-specified crossover probability pc".

After new strings are generated, the mutation operation will be taken for replacing the current string with the pre-specified mutation probability pm.

4.5. Crossover Operation

Two parents deliver two posterity. Quite possibly the chromosomes of the two guardians are replicated unmodified as posterity. Quite possibly the chromosomes of the two guardians are arbitrarily recombined (hybrid) to frame posterity. For the most part the possibility of hybrid is in the vicinity of 0.6 and 1.0.

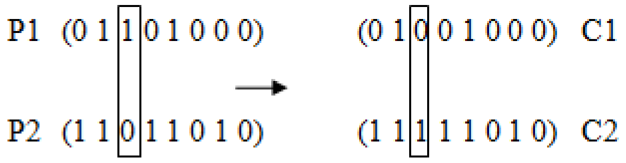

Fig. 4: Example of Crossover

The Fig.4 shows the single point crossover operation. The third bit of the parent strings is crossover to produce new child strings. The main goal of the cross over is to decompose two distinct solutions and then randomly mixes their parts to form novel solutions.

4.6. Mutation Operation

There is a chance that a gene of a child is changed randomly. Generally the chance of mutation is low (e.g. 0.001).

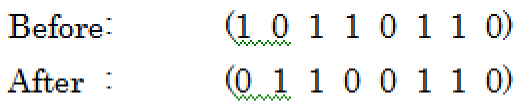

Fig 5: Example of Mutation

In Fig.5 the original binary string is mutated by switching few bits from 1 to 0 or from 0 to 1. The "selection, crossover and mutation are iterated until a pre-specified number Nreplace of new strings are generated".

4.7. Algorithm Summary

The Fig.6 describes the flow of processes involved in the classification system. Initially, a set of rules are generated from the weighted training patterns and then an iterative process is taken for checking fitness and genetic operations until the stopping condition is satisfied.

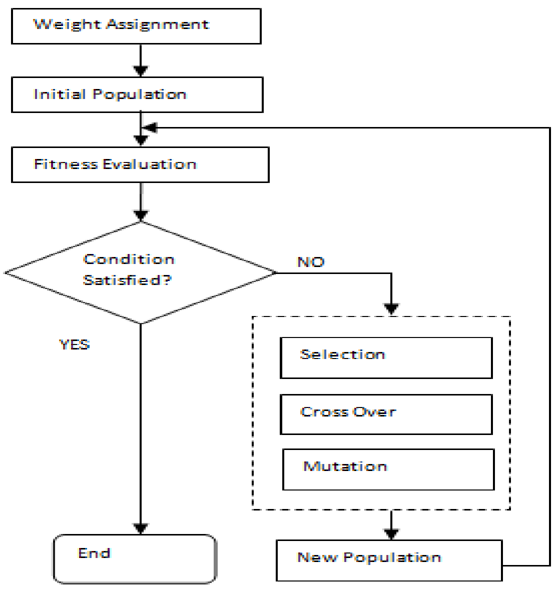

Fig. 6: Flow Chart

The steps involved in hybrid weighted fuzzy rule based classifier are following:
Step 1: "Parameter Specification : Specify the number of fuzzy if-then rules Nrule, the number of replaced rules Nreplace, the crossover probability pc, the mutation probability pm, weights for training patterns wp, and the stopping condition
Step 2: Weight assignment: Assign weights (wp) for each training pattern
Step 3: Initialize: Randomly generate Nrules from weighted training patterns as an initial population
Step 4: Check Fitness: Evaluate the fitness of each rule in the current population, and test if the end condition is satisfied, stop, return the best solution, and exit
Step 5: Genetic Operations: Create a new population based on current population by following genetic operations
Selection: Select two rules randomly from the current population according to their fitness (the better fitness, the bigger chance to be selected)

4.8. Cross Over

With a crossover probability cross over the selected parent rules to form new offspring (children/rules). If no crossover was performed, offspring is the exact copy of parents

4.9. Mutation

With a mutation probability mutate new offspring at each locus
Step 6: Replace: Place current population using new generated offspring population for a further run of the algorithm
Step 7: Repeat: Go to step 2".

4.10. Fitness Evolution

The fitness value of each run is assessed by grouping all the given preparing designs utilizing the arrangement of fluffy if-then standards in the present populace. The wellness estimation of the fluffy if-then control is assessed by the accompanying wellness work:

fitness (Rj) = NCP(Rj) – wp . NMP(Rj)

where, "NCP(Rj) denotes the number of correctly classified training patterns by rules Rj and NMP(Rj) is the number of misclassified training patterns".

4.11. Cost and Accuracy Evaluation

The cost of misclassification or rejection of fuzzy classification system is calculated as,

$$Cost(S) = \sum_{p=1}^{m} w_p \,.\, z_p(S)$$

Where, "Cost(S) is the cost of misclassification or rejection made by a fuzzy classification system S, m is the number of training patterns, wp is the weight of the training pattern yp, and zp(S) is a binary variable that is determined according to the classification result of the training pattern xp by S: zp(S) = 0 if yp is correctly classified by S and zp(S)=1 otherwise[7]".
The accuracy of the hybrid weighted fuzzy rule based classifier is calculated as,
Accuracy, a=t/n
Where,
t-> number of samples correctly classified.
n-> total number of sample cases.

5. Experimental Results

We have developed a hybrid weighted fuzzy rule based system trained by genetic algorithm. The colon cancer and breast cancer datasets are taken for experimental process. The number of classes used is 2. The colon dataset contain 6 attributes and 400 patterns. The breast cancer dataset contains 523 patters and 9 attributes. For result, we have used the yes or no classes. It is clear that the method with weights for training patterns can obtain better classification and fewer rules. The table 1 shows the parameters of genetic algorithm used for the hybrid weighted fuzzy classifier.

Table 1: Parameter Specification

Parameter	Value
Number of rules, Nrule	20
Crossover probability, Pc	0.9
Mutation probability, Pm	0.1
Number of replaced rules, Nreplace	4
Stopping Condition Cycle	50

6. Conclusion

In this paper, a hybrid weighted fuzzy rule based classification system is developed with the combination of assigning weights for training patterns, fuzzy rule generation procedure and genetic algorithm. The Michigan based procedure is used to construct the structure. The compactness of classifier also good compared with existing system. The experimental result shows the hybrid weighted fuzzy rule based classifier gives higher performance than the conventional methods.

References

[1] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," Science, vol. 286, pp. 531–537, 1999.
[2] Pascale Anderle, Manuel Duval, Sorin Draghici, Alexander Kuklin, Timothy G. Littlejohn, Juan F.Medrano, David Vilanova, and Matthew Alan Roberts "Gene Expression Databases and Data Mining" Biotechniques. 2003 Mar; Suppl: 36-44.
[3] Gerald Schaefer, Tomoharu Nakashima, "Data Mining of Gene Expression Data by Fuzzy and Hybrid Fuzzy Methods" IEEE Trans. Information Technology in Biomedicine, vol. 14, pp. 23-29, 2010.
[4] H. Ishibuchi and T. Nakashima, "Improving the performance of fuzzy classifier systems for pattern classification problems with continuous attributes," IEEE Trans. Ind. Electron., vol. 46, no. 6, pp. 1057–1068, Dec. 1999.
[5] Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", 2nd edition, ISBN: 978-55860-901-3, Elsevier.
[6] Max Bramer, "Principles of Data Mining", ISBN: 978-81-8489-166-9, springer, 2007.
[7] Tomoharu Nakashima, Yasuyuki Yokota, Hisao Ishibuchi, "Constructing fuzzy classification systems from weighted training patterns", in proc. 19th European conf. on modeling and simulation, vol 3, pp.2386-2391, 2004.
[8] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," IEEE Trans. Fuzzy Syst., vol. 9, no. 4,pp. 506–515, Aug. 2001.
[9] P.Woolf and Y.Wang, "A fuzzy logic approach to analyzing gene expression data," Physiol. Genomics, vol. 3, pp. 9–15, 2000.
[10] C. Z. Janikow, "A genetic algorithm for optimizing fuzzy decision trees," in Proc. 6th Int. Conf. Genetic Algorithms, Univ. Pittsburgh, Pittsburgh, PA, July 15–19, 1995, pp. 421–428.
[11] S.Sheeba Rani, R.Maheswari, V.Gomathy and P.Sharmila, "Iot driven vehicle license plate extraction approach" in International Journal of Engineering and Technology(IJET) , Volume.7, 2018, pp 457-459, April 2018
[12] M.A.Lee, H.Takagi, "Dynamic control of genetic algorithms using fuzzy logic techniques," in Pmc.Int.Conf. Genetic Algorithm,Urbana-Champaign,lL,July 1993,pp.76-83.
[13] Zhun-Ga Liu, Quan Pan, Jean Dezert, "Hybrid Classification System for Uncertain Data", IEEE Transactions on Systems, Man, and Cybernetics: Systems (Volume: 47, Issue: 10, Oct. 2017).
[14] Balakrishnan S, K.Aravind, A. Jebaraj Ratnakumar, "A Novel

Approach for Tumor Image Set Classification Based On Multi-Manifold Deep Metric Learning", International Journal of Pure and Applied Mathematics, Vol. 119, No. 10c, 2018, pp. 553-562.
[15] A. Jebaraj Rathnakumar, S.Balakrishnan, "Machine Learning based Grape Leaf Disease Detection", Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 08-Special Issue, 2018. Pp. 775-780.
[16] A. Jebaraj Ratnakumar, S. Balakrishnan, S.Sheeba Rani, V.Gomathi, "A Machine Learning Based IOT Device for E-Health Monitoring In a Cloud Environment", Invest Clin. Vol. 58, issue 3, pp. 287-299, 2017. (Web of Science).
[17] S. Vasu, A.K. Puneeth Kumar, T. Sujeeth, Dr.S. Balakrishnan, "A Machine Learning Based Approach for Computer Security", Jour of Adv Research in Dynamical & Control Systems. Vol.10, 11-Special issue, 2018, pp. 915- 919.
[18] Balakrishnan, S., Janet, J., Sujatha, K., & Rani, S. (2018). An Efficient and Complete Automatic System for Detecting Lung Module. Indian Journal Of Science And Technology, 11(26). doi:10.17485/ijst/2018/v11i26/130559