A ROUGH SET BASED FUZZY INFERENCE SYSTEM FOR MINING TEMPORAL MEDICAL DATABASES

U Keerthika[1] R Sethukkarasi[2] and A Kannan[3]

[1] PG Student, Department of Computer Science and Engineering, R.M.K. Engineering College, Kavaraipettai, Tamil Nadu, India
keerthi.umapathy@gmail.com

[2] Research Scholar, Department of Information Science and Technology, Anna University, Chennai, Tamil Nadu, India
sethumaaran@yahoo.co.in

[3] Professor, Department of Information Science and Technology, Anna University, Chennai, Tamil Nadu, India
kannan@annauniv.edu

ABSTRACT

The main objective of this research work is to construct a Fuzzy Temporal Rule Based Classifier that uses fuzzy rough set and temporal logic in order to mine temporal patterns in medical databases. The lower approximation concepts and fuzzy decision table with the fuzzy features are used to obtain fuzzy decision classes for building the classifier. The goals are pre-processing for feature selection, construction of classifier, and rule induction based on increment rough set approach. The features are selected using Hybrid Genetic Algorithm. Moreover the elementary sets are obtained from lower approximations are categorized into the decision classes. Based on the decision classes a discernibility vector is constructed to define the temporal consistency degree among the objects. Now the Rule Based Classifier is transformed into a temporal rule based fuzzy inference system by incorporating the Allen's temporal algebra to induce rules. It is proposed to use incremental rough set to update rule induction in dynamic databases. Ultimately these rules are categorized as rules with range values to perform prediction effectively. The efficiency of the approach is compared with other classifiers in order to assess the accuracy of the fuzzy temporal rule based classifier. Experiments have been carried out on the diabetic dataset and the simulation results obtained prove that the proposed temporal rule-based classifier on clinical diabetic dataset stays as an evidence for predicting the severity of the disease and precision in decision support system.

KEYWORDS

Fuzzy Rough Sets, Lower approximations, Rule Based Classifier, Allen's Temporal Algebra

1. INTRODUCTION

Medical Data Mining is the study of facts in medicine which change with respect to time which begins with the construction of temporal clinical databases. The study helps to generate distinct medical models, in order to foresee a patient's physical condition or recommend medical remedy. The uncertain data can be well handled by the use of fuzzy rough sets, since fuzzy sets take values 0 or 1 only to indicate the degree of trueness of a hypothesis for a given time [9]. These sets are described using a pair of approximations known as upper and lower approximations. The lower approximation of a set refers to the elements that certainly fit into the set, and the upper approximation of a set refers to the elements that perhaps incorporate to the set

in [1].In this paper, a clinical data mining algorithm that uses fuzzy temporal rough sets has been proposed. In this work the problem is well defined as detailed below. Let U be a universe which is the collection of finite set of clinical records. A fuzzy rough set of the given set of clinical records have been formed with fuzzy approximation operations. Let $P_i = \{p_1, p_2, \ldots, p_n\}$ be the set of clinical records and $A_i = \{a_1, a_2, \ldots, a_n\}$ are the set of condition attributes at the time t_i then the lower approximation operation of a fuzzy rough set is applied to construct a temporal fuzzy rough set $\{\underline{S}X, P_i, A_i, t_i\}$ to address the degree of uncertainty.

The lower approximation operation for temporal fuzzy rough set in the approximation space S of the set X is defined based upon the definition given in [1],

$$\underline{S}X = \{ x_i \in U \mid x_i \subset X \} \forall t_{start} < t_i < t_{end} \qquad (1.1.1)$$

where x_i is the set of objects that belongs to X.

The rule is defined as Rule: (Condition) $\rightarrow$ y, where Condition is a conjunction of attributes and y is the class label. Fuzzy Rule Based Classifier is used to categorize the records [7] based on some criteria by means of linguistic values. These linguistic values are identified by certain membership functions. In this   paper temporal fuzzy rule based classifier is build using fuzzy temporal rough sets to make decisions in temporal clinical databases. Allen's temporal algebra offers a composition table that can be used to induce rules. At last prediction is performed to determine the severity of the diabetes disease and to provide medical solution for the patient
This paper is organized in the following manner, In Section 2 we have discussed the related work. We define the proposed system and a few related terms in section 3, 4, and 5. We indicate our conclusion and future work in the section 6.

2. RELATED WORK

B.Walczak *et al* [1] The uncertain data can be well handled by the use of fuzzy rough sets, since fuzzy sets are sets in which elements have degrees of membership, they allow the gradation of truth values. The fuzzy set theory permits the steady evaluation of the membership of temporal data elements in a set; where the operations can be explained with the support of a membership function valued in the interval [0, 1]. These sets are described using a pair of approximations known as upper and lower approximations using the constructive and axiomatic approach. The lower approximation of a set refers to the elements that certainly fit into the set, and the upper approximation of a set refers to the elements that perhaps incorporate to the set in The constructive approach is used to define the lower and upper approximation operations in the form of binary relations, partitions of universe, and Boolean sub algebra. In the axiomatic approach, the approximation operations are defined by using axioms.

D.S.Yeung *et al* [2] provides the concepts of fuzzy rough sets by using the constructive and the axiomatic approaches. From the viewpoint of constructive approach, proposed a work on the generalization of fuzzy rough sets in order to narrate the definitions of upper and lower approximation operators in fuzzy sets by means of arbitrary fuzzy relations to relate the associations between special fuzzy relations and upper and lower approximation operators of fuzzy sets. In axiomatic approach, the different classes of generalized upper and lower approximation operators of fuzzy sets are defined by different sets of axioms. The drawbacks are upper & lower approximation operators of fuzzy sets are defined by arbitrary relations and the axiomatic approaches have been given less importance and the knowledge discovery methods of fuzzy systems are not focused. .

E.C.C. Tsang *et.al* [3] introduced a rule induction technique based on fuzzy rough set. The consistence degree is projected as the vital value to induce rules. A discernibility array is constructed based on the consistence degree, and then an algorithm to find the reduct rule using the discernibility array is designed. The negative aspect is that rule inducing method should be enhanced to reduce the size of the minimal rule set and improve the classification precision.

Wojciech Froelich *et.al* [5] proposed Fuzzy cognitive maps (FCMs) with temporal concepts that are used to find out temporal reliance [5] among medical concepts. The   concepts namely the medical intercession and health issues are stated by variation in patients' conditions. The limitation is FCMs should be analyzed more for temporal concepts.

X.L. Bing *et al* [6] proposed a new approach as a modification to a standard fuzzy modeling method based on the table look – up scheme to foresee disordered time series. This new approach obtains arbitrary values by considering the statistical properties of training dataset. The experimentation was done based the Mackey-Glass time series. It was demonstrated that the modified table look-up scheme can predict the time series more accurately with small number of membership functions even when the noise was added to time series. The issue is that by using large number of membership functions, the method can not predict more accurately.

Suyun Zhao *et. al* [7] put forward a rule-based classifier which is constructed through generalization of Fuzzy Rough Set (FRS) by adhering to a new notion called as consistence degree it is employed as the decisive value to   retain the discernibility information similarly as it is in the procedure of rule induction. The main concern is that the strictures such as threshold value and noise percentage in GFRS are not elucidated in depth.

Giovanni *et. al* [8] presented a idea that Fuzzy Cognitive Maps (FCMs) are used to model and correspond to the behaviour of easy and composite systems by annexing and emulating the human being to depict systems characterized by tolerance, vagueness and granulation of information. They proposed TAFCM (Timed Automated FCM) to work on temporal uncertainty, but evolutionary computation algorithms such as genetic algorithms are not considered. So, the fuzzy temporal rough sets in combination with the genetic algorithm are used in pre-processing the data.

E.C.C.Tsang *et al* [13] focuses on the attributes reduction with fuzzy rough sets. After scrutinizing the preceding works on attributes reduction with fuzzy rough sets, the formal concepts of attributes reduction with fuzzy rough sets are introduced and completely study the structure of attributes reduction. The attributes reduction with fuzzy rough sets is delineated and analyzes its structure in detail. After various discussions on the previous work of reduction with fuzzy rough sets, the formal definition of reduction with fuzzy rough sets is described. Then the structure of the proposed reduction with fuzzy rough sets is analyzed and that all the reductions can be obtained by using the method of discernibility matrix. An algorithm utilizing discernibility matrix to compute all the attributes reductions is developed. The disadvantages are different degrees of similarity cause information loss and algorithm to compute reduction is slow.

II-Seok Oh *et al* [17] proposed an unique hybrid genetic algorithm based approach for feature selection. The hybridization technique produces important effects such as a considerable improvement in the final performance and the acquisition of subset-size control. The hybrid GAs showed better performance when compared to the conventional GAs. A technique of performing rigorous timing analysis was developed, in order to compare the timing requirement of the standard and the proposed algorithms. Experiments performed with various standard data sets revealed that the proposed hybrid GA is superior to both a simple GA and sequential search algorithms. The issues to be concentrated are gene rearrangement for chromosomal encoding and more apposite genetic operators.

The main issue concentrated in this paper is that the lower approximation operation - threshold values have been given more significance in order to intensify the Temporal Rule Based Classifier, thus obtaining firm rule set and from those rule set, the rules are induced by means of the Allen's temporal algebra and prediction is performed to achieve the desired results.

3. PROPOSED WORK

This section describes the proposed system namely, the knowledge base, rule generation subsystem, and prediction subsystem.

To construct the temporal decision fuzzy table for classification, data pre-processing is first performed. We have taken diabetic dataset which is defined with 17 attributes [5]. We propose a genetic algorithm based algorithm for attribute subset selection in order to build a better temporal fuzzy approximation space to remove the obsolete data from the dataset. In order to build temporal rule based classifier using generalization of temporal fuzzy rough set approach with the set of knowledge base is build from clinical temporal databases. The parameters such as threshold α and noise percentage β are taken into consideration in turn for building an efficient rule based classifier to classify the records effectively. The fuzzification process is performed on the crisp values obtained from the previous module. Discernibility vector is designed to describe the consistency degree between the two clinical records. The rules (patterns) are generated by transforming the rule based classifier to fuzzy inference system using the reduced temporal fuzzy decision table which consists of attribute values at a particular time interval have been interpreted as patterns, stated as temporal rules by Allen's temporal algebra to define relations between the temporal intervals. From the rules induced prediction process is performed to conclude the patient's condition pertaining to the diabetes disease. It asserts whether the patient been diagnosed with suffering from diabetes or not, if so to analyze the severity of the disease and plan the medical treatment for the disease.
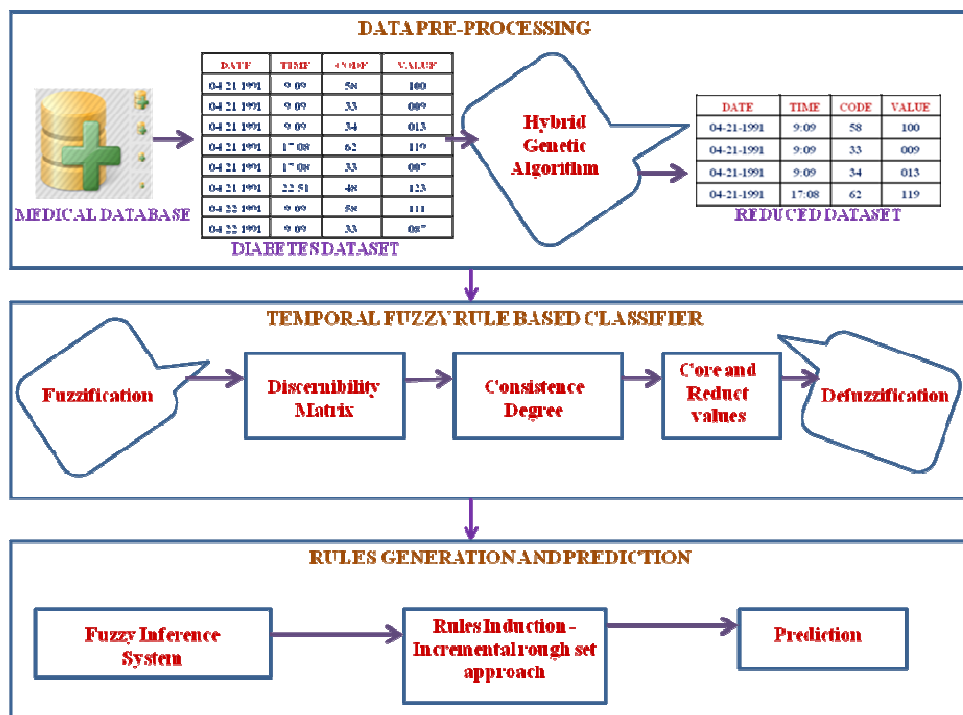

Figure 1: System Architecture

4. METHODOLOGY

4.1 Data Pre-processing

Data pre-processing is generally performed to find out if there is much extraneous and redundant information present on the dataset. This includes various steps such as cleaning normalization, transformation, feature extraction and selection, etc. The outcome of data pre-processing is the ultimate data set with reduced attributes. Feature selection is the technique of selecting a subset of relevant features for constructing best learning models. Genetic algorithms are used for feature selection. Genetic Algorithm is a search algorithm which provides a way for solving problems by imitating the same process that Mother Nature uses. We use the concept of Hybrid Genetic algorithm to determine the reduced attributes. They use the same combination of selection, crossover and mutation to develop a solution to a problem.

The hybrid genetic algorithm concept is explained as to generate N chromosomes which contain a fixed set of variables called genes improved from local operations using roulette wheel selection [17]. It forms the initial population, which includes a group of possible solutions. Compute a fitness value for each chromosome C. It is calculated using the formula as in [17]

$$\text{Fitness(Chromosome)} = J(X_C) - \text{penalty}(X_C) \qquad\qquad (4.1.1)$$

where

X_C is the corresponding feature subset of C, and $\text{penalty}(X_C) = w * ||X_C| - d|$ with a penalty coefficient w, size value d.
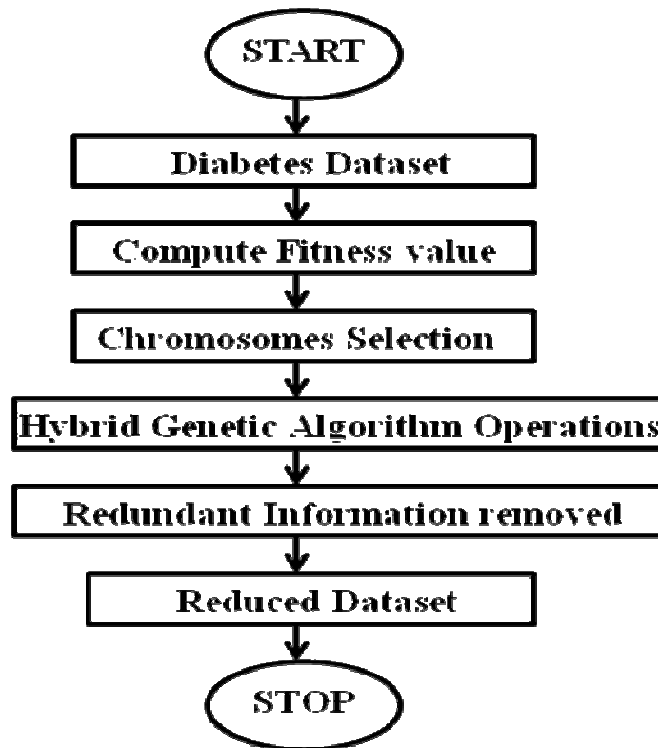

Figure 2: Flow Chart for Hybrid Genetic Algorithm

Using fitness values rank the chromosomes. Select the chromosomes for next generation by replacing all chromosomes of the old population by the chromosomes selected using fitness values. Add the selected children attributes in the population pool. Compute crossover and mutation operation and again add it to the pool. Repeat the process for the required number of attributes is generated. If any feature repeats itself, and then its copies are eliminated. Once the desired output is met, the process halts. The solution is obtained.

Hybrid Genetic Algorithm for Pre-processing

Input: Diabetic Dataset which consists of patients' medical records.
Output: Reduced Dataset with condition attributes.

Begin
1. Load the patients' dataset from the database that fits in the memory.
2. Identify the temporal attributes from the datasets.
3. The dataset is converted to relational database.
4. Chromosomes are selected using roulette wheel method and rank it using fitness value.
5. Eliminate the redundant attributes by applying the hybrid genetic algorithms operations such as crossover and mutation.
6. Selection of attributes is done to measure the consistence degree between the two attributes, those are called as condition attributes.
End

In this work pre-processing is carried out on the diabetes dataset which consists of patient's diabetes history. The dataset has the date, time, code, and value of insulin levels. The insulin levels of patients show a discrepant value according to time. The disease is witnessed by metabolic outcome; the most important being the high blood glucose level, it can be recognized by various blood glucose dimensions. The purpose of the medication is to bring down the standard blood glucose level of the patient to the usual range. The frequency of injections and blood glucose dimensions are indispensable for the patient. Each check up of patient's glucose level is presented in the subsequent order, i.e., date, time (hour and minute), code, and value. There are three tags to recognize the disparate kinds of insulin dosages, i.e., 33, 34, and 35. The other existing measurements are defined in [11].

4.2 Construction of Temporal Fuzzy Rule Based Classifier (TFRBC)

The Temporal Rule Based Classifier is constructed using Fuzzy Temporal Rough Set approach. Generalization of FRS is done using lower approximation operators. α is threshold value applied in lower approximation operation. FRS uses consistence measure which sustains the discernible value invariant. The parameters such as threshold value α used in the approximation operators and noise percentageβ, in addition to other parameters are used to intensify the classifier. Fuzzy decision table is a decision table with fuzzy attributes. The simple lower approximation operators of a fuzzy set is defined in [7] for every $A \in F(U)$, $F(U)$ is the fuzzy power set

$$\underline{R}(A)(x) = \inf{}_{u \in U} \vartheta \, (R(x,u), A_t(u))$$
$$\forall \, t_{start} < t < t_{end} \tag{4.2.1}$$

where ϑ is the residuated implicator, R describes a fuzzy similarity relation and $A_t(u)$ corresponds to activation function. The t_{start} and t_{end} defines the starting time and the ending time of the temporal concept. The lower approximation operation using generalization of fuzzy rough sets is

$$\underline{R}_{\vartheta\alpha}(A)(x) = \inf_{\substack{A_t(u) \le \alpha}} \vartheta(R(x,u), \alpha)$$
$$\wedge \inf_{\substack{A_t(u) > \alpha}} \vartheta(R(x,u), A_t(u)), \forall x \in U$$
$$\forall\, t_{start} < t < t_{end} \qquad (4.2.2)$$

delineated in [7] using the triangular conorm (S) and fuzzy relation. The temporal concepts are included in the activation function $A_t(u)$, where ϑ is the residuated implicator, R describes a fuzzy similarity relation, $A_t(u)$ corresponds to activation function and threshold value α.

$$\underline{R}_{\vartheta\alpha}(A)(x) = \inf_{\substack{A(u) \le \alpha}} S(N(R(x,u)), \alpha)$$
$$\wedge \inf_{\substack{A_t(u) > \alpha}} S(N(R(x,u), A_t(u)), \forall x \in U$$
$$\forall\, t_{start} < t < t_{end} \qquad (4.2.3)$$

The lower approximation hypothesis and fuzzy decision table are used to obtain fuzzy decision classes for constructing the classifier. The lower approximations are used to define the decision classes. By taking a subset of attributes into consideration, the lower approximations are designed. The elementary sets will be the outcome of this stage from which the objects are classified into the decision classes. The decision classes are fixed as 0 and 1. The value 0 and 1 indicates that a patient belongs to category that he/she is and not suffering from diabetes. The fuzzification process is performed for the values on the crisp values to obtain the fuzzy values using the Gaussian membership function, which is defined as

$$e^{-\frac{1}{2}}[(y - c)/\sigma] \qquad (4.2.4)$$

where y represents the input value.

c represents center value of the membership function.

σ represents the difference between the highest and smallest value of the membership function.

4.3 Rules Generation and Prediction

The Rule Based Classifier is transformed into a rule based fuzzy inference system by incorporating neuro fuzzy rules with Allen's temporal algebra to get if then fuzzy rules. The rule is defined as rule: (Condition) $\rightarrow$ y, where Condition is a conjunction of attributes and y is the class label. For reasoning about the relations between temporal intervals, Allen's Interval Algebra provides a composition table which defines probable associations between time intervals that can be used as a basis for reasoning about temporal descriptions of events.

Slot	Starting time (t_{start})	Time(t)	Ending time (t_{end})
1	0	$\le t \le$	6
2	7	$< t \le$	12
3	13	$< t \le$	18
4	19	$< t <$	24

Table 1: Time slots defined for a day

We are taking the time slot in to consideration to add these time values along with the insulin doses. The time in day is divided as six hours for each slot.Allen's thirteen fundamental relations illustrate all the potential relations that the two definite intervals can have.

A reduct is delineated as a nominal necessary subset of a set of attributes that provides the same knack to discern concepts as the full set of attributes is used. The reducts represent necessary condition attributes to make a decision. In a dynamic database if a new object is send in to existing rule set, the existing method need not be modified; instead the increment rough set approach is used to induce the rules which is described in the algorithm [18] given below.

Algorithm for Rule Induction

Input   : Patterns (rules) induced by applying rule based classifier.
Output : Reduced rule set.

Begin
  1. Generate Temporal Fuzzy Rule Based Classifier
  2. Add new data set into raw dataset
  3. Compute the reducts of new dataset.
  4. Calculate Strength Index of these reducts
  5. The reduct with maximum Strength Index from each object is the final rule of this object
  6. Update the final REA rule
  7. Output Final rule sets
End

From the discernibility vector, the core values are defined to find the reducts. The reducts in combination with the association rule mining are used to define the frequent temporal patterns. From these temporal patterns the rules are defined in accordance to the reduct values. The reducts are the reduced attribute values, and then the covering degree [7] of the rule is developed in consideration with the time slots, if the reduct values has the highest covering degree value then those rules are considered as reduced ruleset. In view that the fuzzy if then rules induced from the preceding part using the Allen's temporal algebra the patient's insulin levels are delineated. Based on these values forecasting is done. The defuzzification process is done to convert the fuzzy values to crisp values. By constructing a Rule Based Classifier the prediction process is performed to conclude the patient's   condition pertaining to the diabetes disease. Based upon Allen's temporal interval algebra, the temporal patterns are written as temporal rules for predicting the severity of the disease.

5. RESULTS AND DISCUSSION

In this part the experimental results are discussed, in the pre-processing stage the inconsistent data is removed. In 70 files each file consist of 42 patients details, after pre-processing the redundant data is removed, following which the patient's details are reduced to 25. The table 2 shows a Fuzzy decision table in which fuzzy attributes and decision values are present. The fuzzified values for patient's insulin values present in table 2 are shown in table 3. The results are presented.

p/a	a₁	a₂	a₃	a₄	dc
p₁	100	009	013	119	1
p₂	007	216	081	147	1
p₃	009	084	014	305	0
p4	154	016	007	123	0
p₅	100	009	013	119	1
p₆	007	216	081	147	0
p₇	003	155	123	009	1
p₈	100	009	013	119	1
p₉	154	016	111	225	0
p₁₀	007	216	081	147	1

Table 2: Fuzzy Decision Table

$P_i = \{p_1, p_2, \ldots, p_n\}$ - set of clinical records at the time t_i
$A_i = \{a_1, a_2, \ldots, a_n\}$ - condition attributes at the time t_i
dc - decision class value (value is 0 or 1)

p/a	a₁	a₂	a₃	a₄	dc
p₁	0.9821	0.8460	0.8515	0.9870	1
p₂	0.8432	0.8418	0.9520	0.9427	1
p₃	0.8460	0.9567	0.8530	0.7276	0
p₄	0.9319	0.8558	0.8432	0.9805	0
p₅	0.9821	0.8460	0.8515	0.9870	1
p₆	0.8432	0.8418	0.9520	0.9427	0
p₇	0.8377	0.9304	0.9805	0.8460	1
p₈	0.9821	0.8460	0.8515	0.9840	1
p₉	0.9319	0.8558	0	0.8295	0
p₁₀	0.8432	0.8418	0.9520	0.9427	1

Table 3: Fuzzy Decision Table with fuzzy values

$P_i = \{p_1, p_2, \ldots, p_n\}$ - set of clinical records at the time t_i
$A_i = \{a_1, a_2, \ldots, a_n\}$ - condition attributes at the time t_i
dc - decision class value (value is 0 or 1)

Now the discernibility vector is constructed for all the five records for the values in table 3, are depicted in table 4, where "p" denotes the patient's records. The criterion for discernibility matrix is that the values of the patient's records of the same decision class cannot be compared; only the values of patient's records from different decision classes can be estimated. The individual values of attribute are compared from both the records only the values that are dissimilar are taken into account to construct the discernibility vector. Likewise the entire patient's records are compared and the values are recorded individually. The values with the brackets represents that there exists a consistency between the two objects. The {} represents the null set, it indicates there is no consistency between those two records. The core defines the single attribute values in the

discernibility vector; which should be read from column wise, which is given in the table 5 below.

p	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
p_1	{}	{}	{a1,a2,a4}	{a1}	{}	{a1,a3,a4}	{}	{}	{a1,a4}	{}
p_2	{}	{}	{a2,a3,a4}	{a1,a2,a3,a4}	{}	{}	{}	{}	{a1,a3,a4}	{}
p_3	{a1,a2,a4}	{a2,a3,a4}	{}	{}	{a1,a2,a4}	{}	{a2,a3}	{a1,a2,a4}	{}	{a2,a3,a4}
p_4	{a1}	{a2,a3,a4}	{a1,a2,a4}	{}	{a1}	{}	{a1,a2,a3,a4}	{a1}	{}	{a1,a3,a4}
p_5	{}	{}	{a1,a2,a4}	{a1}	{}	{a1,a3,a4}	{}	{}	{a1,a4}	{}
p_6	{a1,a2,a3,a4}	{}	{}	{}	{a1,a3,a4}	{}	{a2,a3,a4}	{a1,a2,a3,a4}	{}	{}
p_7	{}	{}	{a2,a3}	{a1,a2,a3,a4}	{}	{a2,a3,a4}	{}	{}	{a1,a2,a3,a4}	{}
p_8	{}	{}	{a1,a2,a4}	{a1}	{}	{a1,a3,a4}	{}	{}	{a1,a2,a4}	{}
p_9	{a1,a4}	{a1,a3,a4}	{}	{}	{a1,a4}	{}	{a1,a2,a3,a4}	{a1,a4}	{}	{a1,a3,a4}
p_{10}	{}	{}	{a2,a3,a4}	{a1,a3,a4}	{}	{}	{}	{}	{a1,a3,a4}	{}

Table 4: Construction of Discernibility Vector

$P_i = \{p_1,p_2,....,p_n\}$ - set of clinical records at the time t_i
$A_i = \{a_1,a_2,....,a_n\}$ - condition attributes at the time t_i

p	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
Core	{a1,0.9821}	{}	{}	{a1,0.9319}	{a1,0.9821}	{}	{}	{a1,0.9821}	{}	{}

Table 5: Computation of Core Value

$P_i = \{p_1,p_2,....,p_n\}$ - set of clinical records at the time t_i
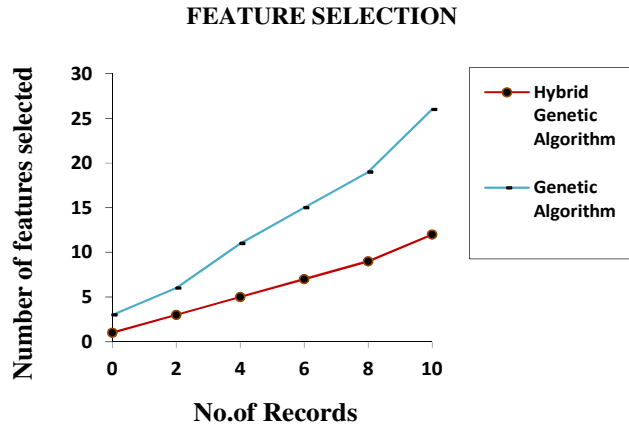
FEATURE SELECTION

Figure 3: Feature Selection – Comparison

In figure 3 the number of features selected is compared with the total number of records. By considering Genetic Algorithms as a feature selection method totally 131 features are selected where while considering Hybrid Genetic Algorithm as the method, 101 features are considered. So in this paper the features are selected using Hybrid Genetic Algorithm.
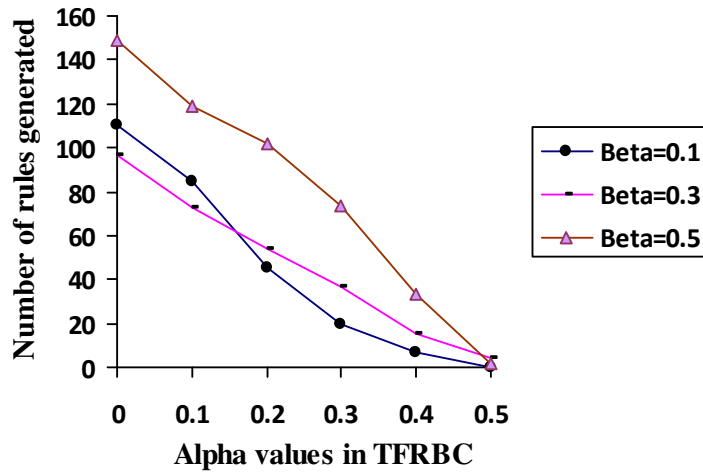
GRAPH WITH ALPHA VALUES

\

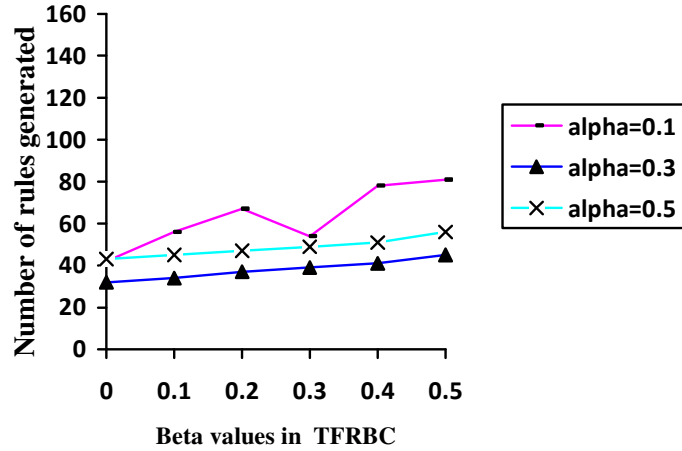Figure 4: Threshold value alpha (α)

GRAPH WITH BETA VALUES

Figure 5: Noise percentage beta (β)

From the computed core value, we can find out the reduct value which is used to work out the final number of rules. The numbers of rules generated against the various threshold values such as alpha and beta (noise percentage) are also compared and their results are discussed. The graph based on the alpha and beta values and number of the rules generated is constructed by varying the beta and alpha values are described in figure 4 and 5 respectively. It shows when β=0.1, the rules generated is around 116; while β=0.2, the number of rule generated is about 99 and β=0.3, the rules obtained is nearly 144. It shows the rules generated is roughly 82; there is also a increase and decrease in the rule generation when α=0.1. By considering the value of α=0.2, the number of rules generated is nearly 52 and while α=0.3, the rules obtained is approximately 41. In figure 6 the accuracies of classifiers such as Fuzzy Set and FRS (Fuzzy Rough Set) are compared. It is shown that accuracy of FRS is more; it is estimated around 52 % when compared with the fuzzy set which is very less nearly 41% when data set value is 600.

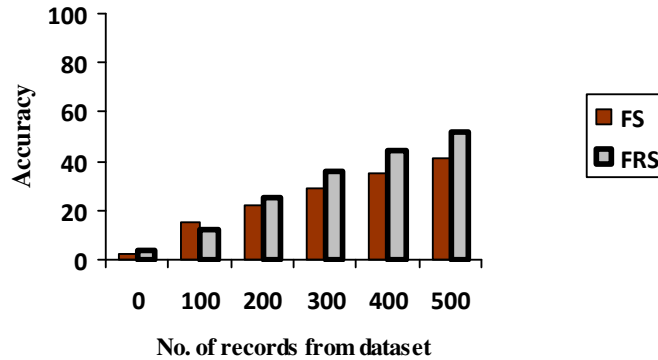**COMPARISON OF ACCURACIES BETWEEN FS and FRS**

Figure 6: Accuracies of FS and FRS

In figure 7 the accuracies of classifiers such as FNN (Fuzzy Neural Network) and TFRBC (Temporal Fuzzy Rule Based Classifier) are compared. It is shown that accuracy of TFRBC is more; it is estimated around 88 % when compared with the FNN which is very less nearly 74%. So the constructed TFRBC is an efficient rule based classifier to effectively classify the clinical records.
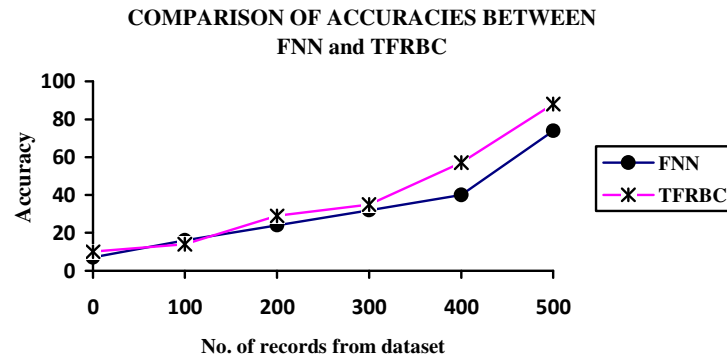
COMPARISON OF ACCURACIES BETWEEN FNN and TFRBC

Figure 7: Accuracies of FNN & TFBRC

6. CONCLUSIONS

The Diabetes dataset is pre-processed using Hybrid Genetic Algorithms so as to eradicate the inconsistent data from the dataset. As Temporal Fuzzy Rule Based Classifier is built by the Generalization of Fuzzy Rough Sets, it's done by the lower and higher approximation operators of similarity relations; the number of human involvement is being decreased and also the newly designed Classifier is simple as compared to alternative classifiers. Within the Rule primarily based Classifier, the Allen's Temporal Algebra is used to induce fuzzy if - then rules. In the Rule Based Classifier, the Allen's Temporal Algebra is utilized to induce fuzzy if - then rules. From the induced rules forecasting is performed to confirm whether the patient is suffering from diabetes or not and to diagnosis the severity of the disease. The analysis of temporal data in medicine is done using the various modules described above. It has begun with the formation of temporal clinical   databases and the efforts to use representations of medical facts for evidential reasoning, which is vital, for developing unswerving medical diagnoses. There is also possibility of using medical models to prophesy a patient's physical condition or devise medical treatment. We have worked and tested our experiments with lower approximation operations alone. In future, upper approximations can be used to build the rule based classifier with effective decisions, which can reduce the number of rule sets when compared with the lower approximations. This helps to build efficient rule based classifier with small number of rule sets.

REFERENCES

[1] B.Walczak and D.L.Massart December 1998, "Rough Sets Theory - Tutorial", Elsevier - Chemometrics and Intelligent Laboratory Systems.
[2] D.S.Yeung, D.G.Chen, E.C.C.Tsang, J.W.T. Lee, and X.Z.Wang, June 2005 "On the Generalization of Fuzzy Rough Sets" IEEE Transactions on Fuzzy Systems, vol.13, pp.343-361
[3] E.C.C.Tsang, S.Y.Zhao, J.W.T.Lee, August 2007 "Rule Induction based on Fuzzy Rough Sets", Proc 2007 Intl Conference on Machine Learning and Cybernetics, vol. 5, pp.3028-3033.
[4] G.Ganesan, D.Latha, and C.Raghavendra Rao, May 2007, "Reduct Generation in Information Systems", Engineering Letters, 14:2, EL_14_2_5 Advance Online Publications.

[5]   Wojciech Froelich, Alicija Wakulicz Deja, May 2009 "Mining Temporal Medical Data Using Adaptive Fuzzy Cognitive Maps" Institute of Computer Science of Silesia, Poland.

[6]   Xinoyu liu Bing, W.Kwan and Simon Y.Foo, 2009 "Times Series Prediction Based on Fuzzy Principles" Department of Electrical and Computer Engineering, Florida State University.

[7]   Suyun Zhao, Eric.C.C.Tsang, Degang Chen, XiZhao Wang, June 2010 "Building a Rule-Based Classifier – A Fuzzy - Rough Set Approach" IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 5, pp.624-638.

[8]   Giovanni Acampora, Vincenzo Loia, June 2011"On the Temporal Granularity in Fuzzy Cognitive Maps" IEEE Transactions on Fuzzy Systems.

[9]   Jyh-Shing, Roger Jang, Chuen-Tsai Sun and Eiji Mizutani, September 1997 "Neuro Fuzzy and Soft Computing – A Computational Approach to Learning and Machine Intelligence"– Prentice Hall of India Private Limited. [10]Timothy Ross, April 2009 "Fuzzy Logic with Engineering and Applications" – Wiley India Second Edition

[11] UCI Machine Repository diabetes dataset http://archive.ics.ucs.edu/ml/datasets/Diabetes.

[12] R.Jensen and Q.Shen, December 2004 "Fuzzy – Rough Sets Assisted Attribute Selection", IEEE Transactions on Fuzzy Systems, vol. 15, pp.73-89.

[13] E.C.C.Tsang, D.G. Chen, D.S.Yeung, X.Z.Wang, J.W.T.Lee, 2008 "Attributes Reduction Using Fuzzy Rough Sets", IEEE Transactions on Fuzzy Systems, vol. 16, pp.1130-1141.

[14] S.Y.Zhao, E.C.C.Tsang, D.G. Chen, April 2009 "The Model of Fuzzy Variable Precision Rough Sets" IEEE Transactions on Fuzzy Systems, vol. 17, no. 2, pp.451-467.

[15] R.Slowinski, D.Vanderpooten, April 2000 "A Generalized Definition of Rough Approximations Based on Similarity" IEEE Transactions on Knowledge and Data Engineering, vol. 12, no. 2, pp.331-336.

[16] Allen.J.F. November 1983 "Maintaining Knowledge about Temporal Intervals." Communications of the ACM 26, 11, 832.

[17] II-Seok Oh, Jin Seon Lee and Byung –Ro Moon November 2004, "Hybrid Genetic Algorithms for feature Selection" IEEE Transactions on attern Analysis and Machien Intelligence, vol.26, no. 11, pp. 1424 – 1437.

[18] Sheng-Uei Guan, and Fangming Zhu April 2005 "An Incremental Approach to Genetic-Algorithms-Based Classification", IEEE Transactions on Systems, Man, and Cybernetics vol. 35, no.2 pp.227-239.

[19] Xizhao Wang, Eric.C.C.Tsang, Suyun Zhao, Degang Chen, Daniel S.Yeung, April 2007 " Learning Fuzzy rules from fuzzy samples based on rough set technique" Information Sciences vol.177, no.20, pp.4493 – 4517.

[20] Ian Cloete and Jacobus van Zyl February 2006 "Fuzzy Rule Induction in a Set Covering Framework", IEEE Transactions on Fuzzy Systems, vol. 14, no.1 pp.93-110.

Authors

Ms. U. Keerthika received her B.E degree in Computer Science and Engineering from Panimalar Engineering College affiliated to Anna University, Chennai. She has completed her post graduation in Computer Science and Engineering at RMK Engineering College affiliated to Anna University. Her area of interest includes Data Mining, Artificial Intelligence and Fuzzy Sets.

Ms. R. Sethukkarasi, graduated in Computer Science & Engineering in the Bharathidasan University., completed PG at RMK Engineering College, Anna University. Currently she is pursuing Ph.D in the Department of Information Science &Technology, Anna University under the guidance of Dr.A.Kannan. Her area of interest includes Data Mining, Temporal Analysis, and Artificial Intelligence.

Dr.A.Kannan, graduated in M.Sc Maths from Annamalai University., completed PG and Ph.D from the Anna University. Currently he is   a professor in the Department of Information Science &Technology, Anna University. His area of interest includes Database Management System, Artificial Intelligence and Software Engineering.