ISSN: 0975-766X
CODEN: IJPTFI
Available through Online ***Review Article***
www.ijptonline.com

HYBRIDIZATION OF SOFT COMPUTING FRAMEWORK-A SURVEY

R.Rathi*,
Assistant professor [Senior], VIT University,Vellore.
Email: rathi.r@vit.ac.in

Received on 15-02-2016 *Accepted on 05-03-2016*

Abstract

Hybridization of different soft computing techniques is very useful in solving real world problems. Soft computing techniques like rough set theory and fuzzy sets are used to handle uncertain or vague information present in the real world data. Hybridizing the soft computing techniques yield better result than the stand alone techniques. Hybridizing these techniques serves different purpose like attribute subset selection, improving classification accuracy, prediction and many more. Data mining concepts are useful in extracting knowledge from the large data set. The usages of some of the soft computing concepts hybridized with data mining concepts used for classification and attribute reduction are also focused and the future work of each concept is highlighted. Comparative analysis made with the survey work is focused which focuses on scalability, robustness, efficiency and accuracy of different classifiers.

Keywords: Hybridization, Rough sets, classification.

1 Introduction

Nowadays, there is a huge rise in the amount of data being used. It is a very hard task to extract knowledge from this huge amount of data. There should be a useful tool to process these data and extract knowledge from that. One such tool is the rough set theory. Rough set theory was discovered by pawlak [1].Rough set extracts knowledge from the huge dataset. The knowledge extracted should be optimized using some powerful optimization algorithm. One such optimization algorithm is genetic algorithm. This survey work starts with the introduction of rough set theory and genetic algorithm as these two topics were major concentrated.

1.1 Introduction to Rough Set Theory

A very powerful mathematical tool proposed by [1] Zdzislaw pawlak [pawlak,(1982)] is rough set theory ,which is used to handle uncertain or vague data present in the information system. In rough set ,the entire dataset is expressed as Information system I=<U,A,V,f> where U denotes the finite set of objects U={x_1,x_2,x_3…x_n}, A denotes the finite

set of attributes A={ $a_1,a_2,a_3 \ldots a_n$ }, V denotes the domain of each attribute A, f denotes the function that defines U×V → Va . An information system which includes a decision attribute is the decision system with A=CUD. Table 1 shows the example decision system. Patients P1,P2,P3,P4,P5,P6,P7 represents an object. Temperature, weakness, headache, nausea represents the condition attribute and flu represents the decision attribute. [3] [Rathi, (2013)]. Rough set is defined with the help of approximations which is lower and upper approximations. Let X be a target set belonging to U (Universal set), We can characterize X by a pair of Upper and Lower approximation. Five different regions of interest are defined

i) Lower approximation of target set X is given by $\underline{RX}= \cup\{Y \in U / R : Y \subseteq X\}$ and it consist of all the members which surely belongs to the set

ii) Upper approximation of target set is given by $\overline{RX} = \cup\{Y \in U / R : Y \cap X \neq \phi\}$ and it consist of all the members which possibly belongs to the set

iii) Negative region (certainly non-member of X) is given by

$NEG_R(X)=U- \overline{RX}$ (1)

iv) Positive region (certainly member of X) $POS_R(X)= \underline{RX}$ (2)

v) Boundary region of X $BN_R(X)= \overline{RX} - \underline{RX}$ We say that *X* is rough with respect to *R* if and only if $\overline{RX} \neq \underline{RX}$, equivalently $BN_R(X) \neq \phi$. *X* is said to be *R*–definable if and only if $\overline{RX} = \underline{RX}$ or $BN_R(X) \neq \phi$. So, a set is rough with respect to *R* if and only if it is not *R*–definable.

Table-1: Decision system.

Patients	Temperature	Headache	Weakness	Nausea	Flu
P1	Very high	Yes	Yes	No	Yes
P2	High	Yes	No	Yes	Yes
P3	Normal	No	No	No	No
P4	Normal	Yes	Yes	Yes	Yes
P5	High	No	Yes	No	Yes
P6	High	No	No	No	No
P7	Normal	No	No	No	Yes

1.2 Introduction to Genetic Algorithm

GA is very powerful search optimization technique discovered by Charles Darwin. Genetic algorithm starts with random population of solution to a problem. From the random population of solution, genetic algorithm finds the best optimum solution. The population of solution or chromosome is evaluated against the fitness function, which is defined for the problem in our hand. The fitness function serves as the solution to the problem. Only the chromosome with the highest fittest value has been selected for the next generation. Crossover operator is applied over the selected chromosome which is used to exchange the properties of one parent with the other parent. Once the crossover is done, mutation operator is applied to the chromosome to get the next new population .The set of procedures starting from fitness evaluation is repeated to get the best optimum solution. The parameters of genetic algorithm are Population size, fitness function, crossover probability, mutation probability. Genetic algorithms performance mainly depends upon the crossover operators and mutation operators.

1.2.1 Chromosome representation

Based upon the problem in our hand chromosome is represented. Chromosome representation takes many forms like Binary representation, permutation representation, tree representation, octal representation[Agarwal(2014)][6].The chromosome for a rule can be represented like "If temperature is high and weakness is no the flu is yes" which can be encoded into chromosome like 20102[Han&kamber,2000][17].

The chromosome length is the number of condition attribute present in the dataset. It can be represented in binary format like 10 for 2 and 01 for 1 and 00 for 0. The general procedure of Hybridizing technique is that classifier is chosen to generate classification rule.

This survey paper is organized as section 2 focuses on literature survey of different hybridization techniques; section 3 shows a table of analysis done in the survey work and its description, section 4 ends with the conclusion.

2 Literature Survey

2.1 Hybridizing rough set theory with genetic algorithms for classification

Hybridizing rough set theory with genetic algorithms produces an optimal solution at low cost so that speed of the process is increased. The output of rough set classifier is given to genetic algorithm for improving classification accuracy. Selection, crossover, mutation are applied successfully to produce new population and that is fed again for further processing by the genetic algorithm. This procedure is repeated until termination criteria is met (i.e.) forecasting accuracy [Cheng, Ching-Hsue, We i,(2010)][18] discussed about the usage of hybridizing rough set and

genetic algorithms for stock price forecasting. They compared the accuracy and stock return of rough set theory, genetic algorithms and hybridization of both rough set theory and genetic algorithm for Taiwan stock exchange dataset. They concluded that hybrid model performs well when compared with other models.

The results with respect to accuracy are shown in table 2 as discussed by Cheng, Ching-Hsue, Tai-Liang Chen, and Liang-Ying Wei .From the table [2], it is clearly understood that Hybrid model produces better accuracy when compared to rough and GA as a standalone model.

Accuracy = Number of rules classified correctly /All of the observations in the training data.

Table-2: Accuracy measure of three models.

Year	Model		
	Roughset theory	Genetic algorithms	Hybrid model
2003	0.512	0.535	0.582
2004	0.571	0.556	0.614
2005	0.488	0.51	0.568

[2] Basabi and goutham chakraborthy have also discussed about finding the right set of rules even after discarded irrelevant attributes. The partitions induced by the different levels of relevant attributes would never lead to rules that are consistent to make decisions. So the right set of rules has to be found out by employing genetic algorithms. The fitness function for finding the right set of rules is max(number of objects classified as decision yes, number of objects classified as decision no)/ (number of objects classified as decision yes+ number of objects classified as decision no).

2.1.1 Future work:

Hybridization of rough set theory with neural networks or fuzzy sets with genetic algorithms or fuzzy sets with neural networks can be done as future work and the resultant can be compared. Scalability feature of the same work can be done by taking large data set. Sensitivity and specificity can also be used to measure the accuracy of the rule.

2.2 Hybridizing Rough set theory with genetic algorithm for feature reduction

The purpose of hybridizing rough set theory with genetic algorithm is to find out the subset of attributes which are relevant to the problem. It is a very tough job to find the important attribute from the dataset.

[Basabi chakraborty and goutham chakraborthy,(2004)][2] discussed about hybridization of rough set and genetic algorithms for selecting subset of attributes. With the help of rough set notations (β defined with respect to condition attributes as well as decision attribute) the best attribute is found. The rough set notation used to find the best attribute

by diminishing the boundary region is the union of positive regions and negative regions which is divided by the count of objects present in the universe [pawlak, (1982)]. Genetic algorithms fitness function is defined over this rough sets β value which is given as fitness function to genetic algorithm for finding out the relevant attributes .Eq 4 shows the fitness function.

$\beta_A(B) = POS_A(B)\ U\ NEG_A(B)/U$ (4).

Basabi chakraborthy and goutham chakraborthy done their experiments for the proposed model on Pima Indian diabetes database. Pima indian diabetes database has 8 condition attributes and one decision attribute. The experimental results shows that out of 8 attributes only 4 attributes are the important attributes. [Zuhtuogullari, Kursat, Novruz Allahverdi, and Nihat Arikan, (2013)][5] developed GARSBS (Genetic algorithm and rough set based feature reduction software) for attribute reduction and the output is given as input to neural networks for classification. GARSBS software developed works well for a very large dataset. Information systems which has large input spaces requires high processing times and memory. To overcome this, GARBS software is used.

2.2.1 Future Work

Discernibility matrix can be used to find out the important attributes and that is fed as fitness function to the genetic algorithm. Robustness of the algorithm can be concentrated.

2.3 Hybridizing Fuzzy set theory with genetic algorithm for feature reduction

[Ephzibah,(2011)][4] discussed about the feature reduction using genetic algorithm and fuzzy set theory. Genetic algorithm is used to find the important set of attributes. The fitness function defined for finding the important attribute is the addition of attribute value.

The output of genetic algorithm is given as input to fuzzy set for effective prediction. The popular pima Indian diabetes dataset is taken as the experimental dataset. Mat lab GA tool is used to compare the efficiency of Fuzzy set with genetic algorithms and without genetic algorithms.

It is concluded that Fuzzy set with Genetic algorithm approach gives better result for feature reduction. Initial 8 condition attributes in pima dataset are reduced to 3 or 4 or 5 using the hybrid approach.

Fitness function =addition of attribute values. (5)

2.3.1 Future Work

Genetic algorithm is used for attribute reduction and fuzzy logics classifications are given as input to genetic algorithm to improve rules accuracy.

2.4 Hybridization of genetic fuzzy and neural networks for prediction

[19] Esmaeil Hadavandi, Hassan Shavandi , Arash Ghanbari proposed a model which is the hybridization of genetic fuzzy and neural networks. The proposed model is applied for stock price prediction. Stepwise regression analysis is used for variable selection and neural networks are used to divide the problem space into sub problem space or clusters. Output of neural networks is fed to genetic fuzzy systems for prediction. The input problem is derived into knowledge base by fuzzy rule based systems. Fuzzy rules are stored into knowledge base. By tuning the database, the accuracy is improved using genetic algorithm.

2.4.1 Future work

The same work can be carried out by considering rough set theory instead of fuzzy sets, since rough set yields good results when compared to fuzzy sets.

2.5 Hybridization of neuro fuzzy and genetic algorithms for classification

[Epzibah,2012][7] proposed a neuro fuzzy model for classification. Genetic algorithm is used here for selecting subset of attributes and neuro fuzzy system is used for classification. The proposed model is experimented with heart disease dataset. Classification accuracy rate is calculated as shown in equation (6)

Classification accuracy=Number of instances classified correctly/Number of training instances --(6)

2.5.1 Future work

Future work can be carried out by hybridizing rough set theory with neural networks and its accuracy is compared with the proposed model in [Epzibah,2012][7]. Scalability and robustness of the proposed method can be concentrated.

2.6 Other hybridizations

Generating rules from the numerical attribute data set is obviously not appropriate in case of information table , that are not exactly identical but almost identical which leads to convert the numerical attribute values into symbolic attributes. Therefore, fuzzy proximity relation is adopted to process the information table. Hybridizing both rough set on fuzzy approximation space [D. P. Acharjya, (2008)][11] and LERS classification system (LEM1,LEM2) can be used for inducing rules.

The same task can be carried out by hybridizing rough set on intuitionistic fuzzy approximation spaces. [Acharjya, D. P, (2012)][12] Bayesian classification is used mainly for prediction. It is not applied directly to information system contains almost indiscernible attribute values. To overcome this, hybridization of uses both rough sets with ordering

rules and Bayesian classification is used. Rough set is used to process the almost indiscernible attribute values and the output of rough sets is given as input to Bayesian classification. [Rajni Jain][13] Hybridization of rough set theory and decision theory to solve the problems of computational overheads is discussed.

The computational overhead is compared for rough sets, ID3 algorithm and hybridization of rough set theory with decision tree. The authors concluded that hybridization of rough set theory with decision tree performs well in terms of accuracy.

3. Comparative Analysis of different Techniques

Table-3 shows the analysis done for only classification purpose and all the methods can be successfully applied for clustering, prediction etc. The input given to each method, its purpose, capable of classification, attribute subset selection, scalability, Robustness, speed and accuracy of the methods were discussed. Scalability issues are hot area of research.

3.1 Scalability

In order to support the massively increasing heterogeneous data need of scalability arises. FP growth algorithm can be used and in fuzzy set adaptive fuzzy neuro fuzzy system can be used which divides the entire dataset into n subsets and prediction value are derived with the test value and all should be combined.

In genetic algorithm, parallel genetic algorithm is used to solve the scalability issues for heterogeneous dataset.

3.2 Speed

Speed of all the classifier refers to the computational cost involved or how efficiently the classifier is able to generate the classification rules. In rough set and fuzzy set, the computational overhead is achieved with the help of map reduce concept.

In genetic algorithm its efficiency is calculated by population size and the number of generations it takes to converge. In neural networks efficiency is calculated by Time spent training the network, simulated annealing. In decision tree, it was measured with the number of leaves in a tree and the error rate of the tree.

3.3 Accuracy

Accuracy of the classifier depends upon the classifiers ability to correctly predict the class label of unseen data. The accuracy of the classifier is verified using testing data.

In general, accuracy is calculated as number of objects classified correctly / number of objects classified correctly number of objects not classified correctly.

Table-3: Comparative analysis of different techniques.

Methods	Input	Purpose	Classification	Removal of redundant value	Scalability	Speed or efficiency	Accuracy
Rough set	Information system I=(U,A)	To deal with uncertain data	YES	Discernability matrix[1][3]	YES[FP growth algorithm][8]	MAP-REDUCE [9]	ΣcardRXi / Σcard $\overline{RX}$ i[1]
Genetic algorithm	Population of chromosome	To find optimum solution	YES	?????$_A$(B) ? ?????$_A$(B)/*U [2]*	YES[PGA][10]	Population size and number of generation it takes to converge.[19]	Accuracy=Number of instances classified correctly/Number of training instances
Fuzzy set	Membership value and the information system	To deal with uncertain data	Yes	Using rough set concept[15]	YES[Adaptive neuro inference fuzzy system][14],Feature selection	MAP-REDUCE[9]	Same as Genetic algorithm
Neural networks	Input of training data as input layer	To learn and to extract knowledge from the data.	YES	Rough set with GA ,Principle component analysis	YES(Feature selection)[16]	Time spent training the network,simulated annealing [17]	Cross vaalidation technique
Decision tree	Data set	Classification	YES	Information gain,gain ratio and gini index	YES [SLIQ,SPRINT]	Number of leaves in a tree and the error rate of the tree[17]	Tree pruning
Rough set on Fuzzy approximation space	Categorical data and α values	Convert categorical data to symbolic data	NO	NO	NO	NO	NO

4. Conclusion

This survey work concentrates mainly on the hybridization concepts of rough set theory with genetic algorithm and neural networks , fuzzy sets with genetic algorithm and neural networks and also neuro fuzzy with genetic algorithm ,genetic fuzzy with neural networks for classification as well as attribute reduction. These hybridization concepts works well for several applications. This integration of soft computing techniques has been successfully applied to solve many real time problems. It is concluded that by hybridizing all the models, the performance is achieved better when compared it with the individual model. The performance of all the hybridization models can be compared to achieve more performance, scalability to support a very large data and robustness for correct and errorless classification of all the hybridization techniques is suggested as the future work. Table shown shows the analysis done with respect to scalability,efficiency,accuracy,purpose,input given etc.

5. References

1. Pawlak, Zdzisław. *Rough sets: Theoretical aspects of reasoning about data*. Vol. 9. Springer Science & Business Media, 2012.
2. Chakraborty, Goutam, and Basabi Chakraborty. "A rough-GA hybrid algorithm for rule extraction from large data." *Computational Intelligence for Measurement Systems and Applications, 2004. CIMSA. 2004 IEEE International Conference on*. IEEE, 2004.

3. Rathi, R., et al. "Rule Extraction by Rough Set Approach." *International Journal of Applied Engineering Research* 8.14 (2013).

4. Ephzibah, E. P. "Cost effective approach on feature selection using genetic algorithms and fuzzy logic for diabetes diagnosis." *arXiv preprint arXiv:1103.0087* (2011).

5. Zuhtuogullari, Kursat, Novruz Allahverdi, and Nihat Arikan. " rough set theory and genetic algorithms based hybrid approach for reduction of input attributes in medical systems" (2013).

6. Aggarwal, Shaifali, Richa Garg, and P. Goswami. "A review paper on different encoding schemes used in genetic algorithms." *International Journal of Advanced Research in Computer Science and Software Engineering* 4.1 (2014): 596-600.

7. Ephzibah, E. P., and V. Sundarapandian. "A neuro fuzzy expert system for heart disease diagnosis." *Computer Science & Engineering: An International Journal (CSEIJ)* 2.1 (2012): 17-23. [17]

8. Patil, Prachi. "Data Mining with Rough Set Using Map-Reduce." *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE) ISSN (Online)* (2014): 2320-9801.

9. Kwiatkowski, Piotr, Sinh Hoa Nguyen, and Hung Son Nguyen. "On Scalability of Rough Set Methods." *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Methods*. Springer Berlin Heidelberg, 2010. 288-297.

10. Liu, Yan Y., and Shaowen Wang. "A scalable parallel genetic algorithm for the generalized assignment problem." *Parallel Computing* (2014).

11. D. P. Acharjya, and B. K. Tripathy, "Rough Sets on Fuzzy Approximation Space and Application to Distributed Knowledge Systems", International Journal of Artificial Intelligence and Soft Computing, vol. 1 (1), (2008), pp. 1-14.

12. Acharjya, D. P., Debasrita Roy, and Md A. Rahaman. "Prediction of missing associations using rough computing and Bayesian classification."*International Journal of Intelligent Systems and Applications (IJISA)* 4.11 (2012): 1.

13. Sonajharia Minz, Rajni Jain "Rough Set Based Decision Tree Model for Classification" Volume 2737 of the series Lecture Notes in Computer Science pp 172-181

14. Lawrence O. Hall University of South Florida, USA "Scalable fuzzy algorithms for Data management and analysis,Chapter 2-Scalable fuzzy models".

15. Tsang, Eric CC, et al. "Attributes reduction using fuzzy rough sets." *Fuzzy Systems, IEEE Transactions on* 16.5 (2008): 1130-1141.

16. Peteiro-Barral, Diego, et al. "Toward the scalability of neural networks through feature selection." *Expert Systems with Applications* 40.8 (2013): 2807-2816.

17. Han, Jiawei, and Micheline Kamber. "Data mining: concepts and techniques (the Morgan Kaufmann Series in data management systems)." (2000).

18. Cheng, Ching-Hsue, Tai-Liang Chen, and Liang-Ying Wei. "A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting." *Information Sciences* 180.9 (2010): 1610-1629.

19. Hadavandi, Esmaeil, Hassan Shavandi, and Arash Ghanbari. "Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting." *Knowledge-Based Systems* 23.8 (2010): 800-808.

Corresponding Author:

R.Rathi*,

Email: *rathi.r@vit.ac.in*