Association Rule Based Pattern Analysis with Liver and Diabetes Dataset

Shiv Shakti Shrivastava[1], Dr. V.K.Choubey[2], Dr. Anjali Sant[3]
[1]Research Scholar, CS, Mewar University, Chittorgarh. (Rajsthan) India.
[2,3]Mewar University, Chittorgarh (Rajsthan) India.

ABSTRACT
Association rule mining is important tool for data analysis. The associations rule mining technique finds the relation between two different attributes on the basis of support and confidence value. The association rule mining technique gives two algorithms for analysis of data, one is memory based and the other is memory independent. The memory based algorithm is well known algorithm called apriori algorithm and memory impendent algorithm is called FP-growth algorithm. The rough set based rule generation technique is a new approach for classification. In this technique rough set generates a number of rules, these generated rules design the class builder and finally data are classified. The rough set based classification technique reduces the number of attributes during the classification process. For the optimization and better prediction of classification gravitational search optimization technique in combination of rough set theory is used. This paper discusses the rough set theory, rule generation algorithm and Gravitational Search Algorithm (GSA), proposed algorithm and proposed model for the classification process.

I ROUGH SET THEORY

The rough sets theory deals with uncertain and fuzzy materials and helps simplification. In the rough sets theory, humans use their general knowledge to classify the world around them as abstract or concrete. Everything is classified according to its characteristics, and those with nearly identical characteristics may be put into the same group. One of the main advantages of rough set theory is that it does not need any preliminary or additional information about data. The main problems that can be approached using rough sets theory includes data reduction, discovery of data dependencies, and estimation of data significance, generation of decision algorithms from data, approximate classification of data, discovery of patterns in data and discovery of cause-effect relationships [10]. The following is the concept of rough sets theory.

II GRAVITATIONAL SEARCH ALGORITHM [GSA]

The Gravitational Search Algorithm (GSA) is simulation of Newton's gravitational force behaviors. In this algorithm, possible solutions of the problem in hand are considered as objects whose performance (quality) is determined by their masses, all these objects attract each other by the gravity force that causes a global movement of the objects towards the objects with heavier masses. The position of each object corresponds to a solution of the problem, and inertial masses are determined by a fitness function. The heavy masses, which represented a good solutions, move more slowly than lighter ones, this represents the exploitation of the algorithm. The GSA starts with a set of agents, selected at random or based on some criteria, with certain positions and masses representing possible solutions to a problem, and iterates by changing the positions based on some values like fitness function, velocity and acceleration that gets updated in every iteration

III PROPOSED METHODOLOGY

The proposed algorithm of rough set based classification technique uses gravitational search algorithm. The rough set based classification technique uses rule generation technique form the process of rough set. The rough set is a combination of fuzzy logic and classical set theory. In process of objective function minimization gravitational search algorithm for the process of rule optimization is used. The process of rule optimization technique reduces the number of rules during the class builder process. The GSA algorithm divides the all rule segment in three class i.e. upper class, lower class and average class. Following are the steps of algorithm-

(a) Data are passes through RGI
(b) RGI gives the number of rules
(c) All rules divide into three section lower upper and average
(d) The training phase data are passes through GSA sampler
(e) The sampling of data passes through class builder balanced the data for minority and majority ratio of class
(f) The sampled data assigned to k-type binary class
(g) Binary class data are coded in bit form
(h) if code bit value is single assigned the class value
(i) Else data goes to training phase
(j) . Balanced part of training is updated
(k) Find accuracy and relative mean Error
(l) Exit

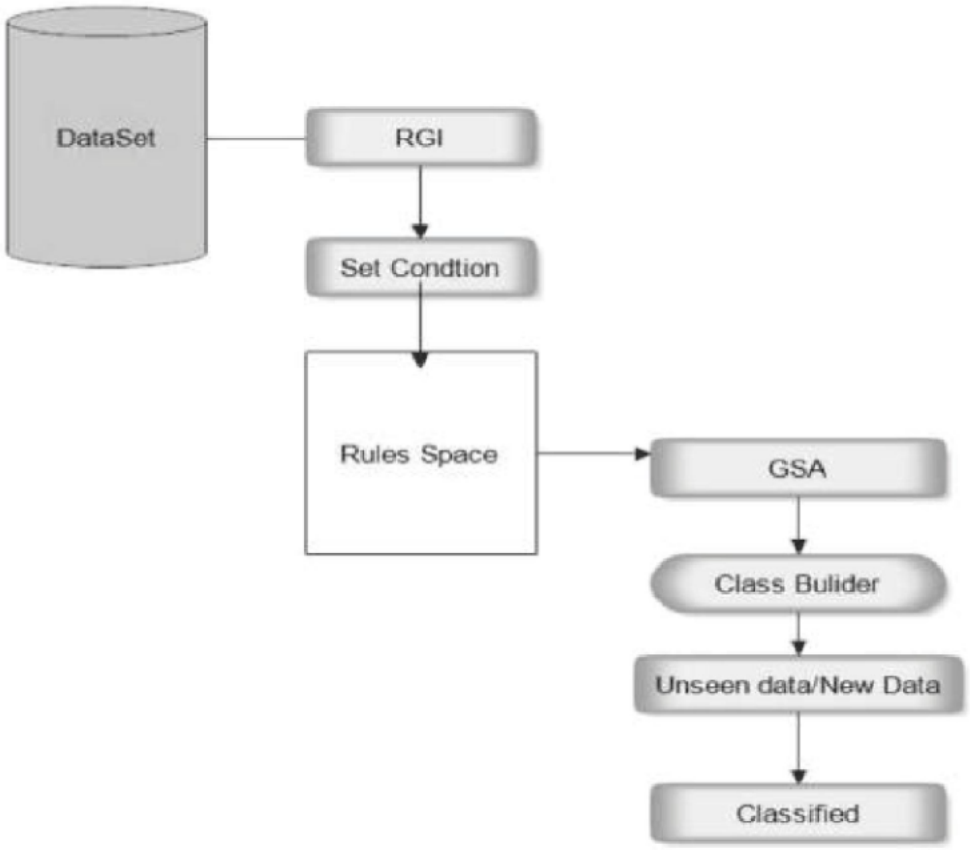

Fig. 1: Proposed models for rule based classification based on rough set and GSA algorithm.

IV EXPERIMENTAL RESULTS AND PROCESS DATASET DESCRIPTION LIVER AND DIABETES DATASET

(a) Attribute information:

(i) mcv mean corpuscular volume
(ii) alkphos alkaline phosphotase
(iii) sgpt alamine aminotransferase
(iv) sgot aspartate aminotransferase
(v) gammagt gamma-glutamy transpeptidase
(vi) drinks number of half-pint equivalents of alcoholic beverages drunk per day
(vii) selector field used to split data into two sets
(viii) Missing values: none

(b) Performance Evaluation and Result Analysis

Table 1
Accuracy and elapsed time tabulation

Data set Name	Method	Support	Confidence	Accuracy (%)	Elapsed Time (sec)
Liver	CBA	0.3	0.5	81	6.245
	RGI	0.3	0.5	82	5.308
	Proposed	0.3	0.5	**86**	**4.633**
Diabetes	CBA	0.3	0.5	81.49	7.245
	RGI	0.3	0.5	83.32	8.451
	Proposed	0.3	0.5	**87.72**	**9.678**

The result tabulated at table 1 shows that the Accuracy and Elapsed time with using CBA, RGI and Proposed techniques for the same and different dataset.

 ISBN: 2278-4187

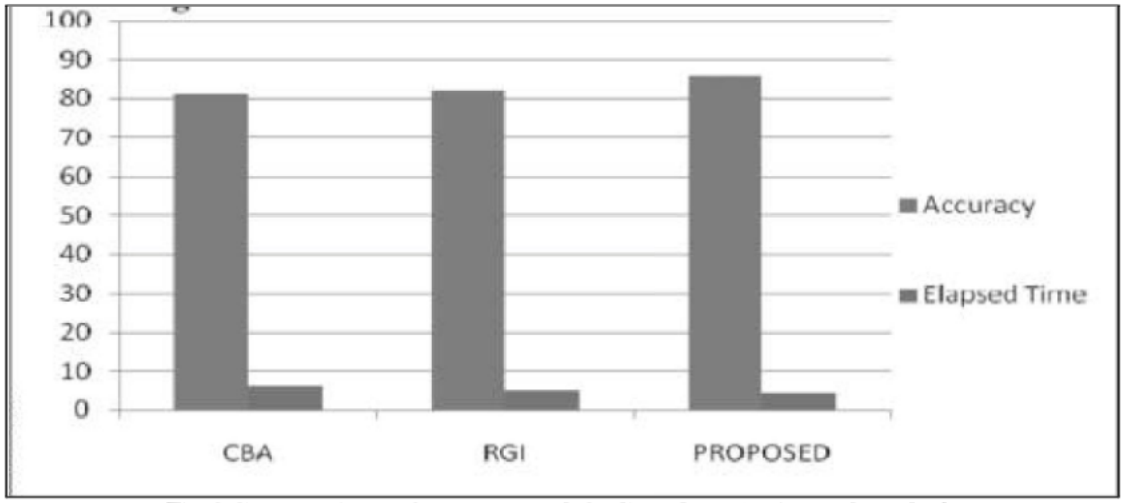

Fig. 2 Comparative performance graph for liver dataset using each method

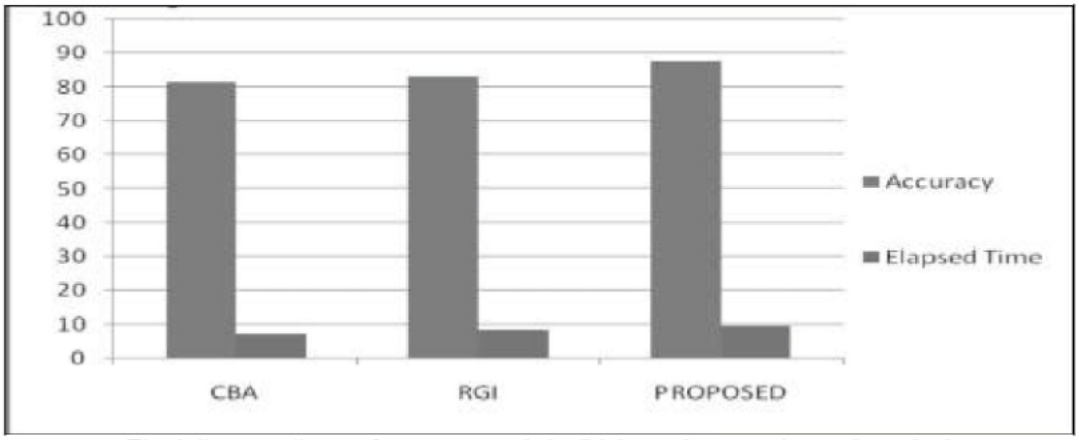

Fig. 3 Comparative performance graph for Diabetes dataset using each method

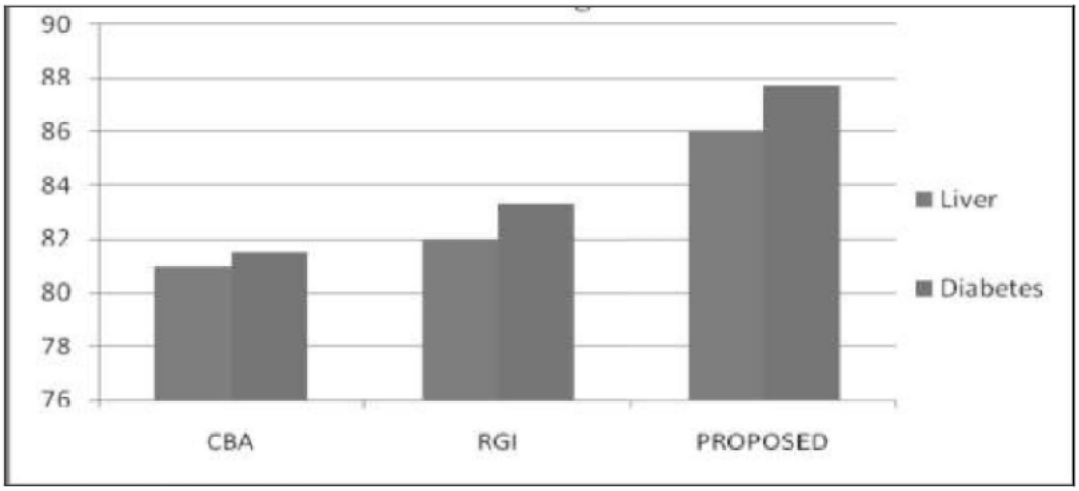

Fig.4 Comparative performance graph for Accuracy using Liver and Diabetes dataset using each method

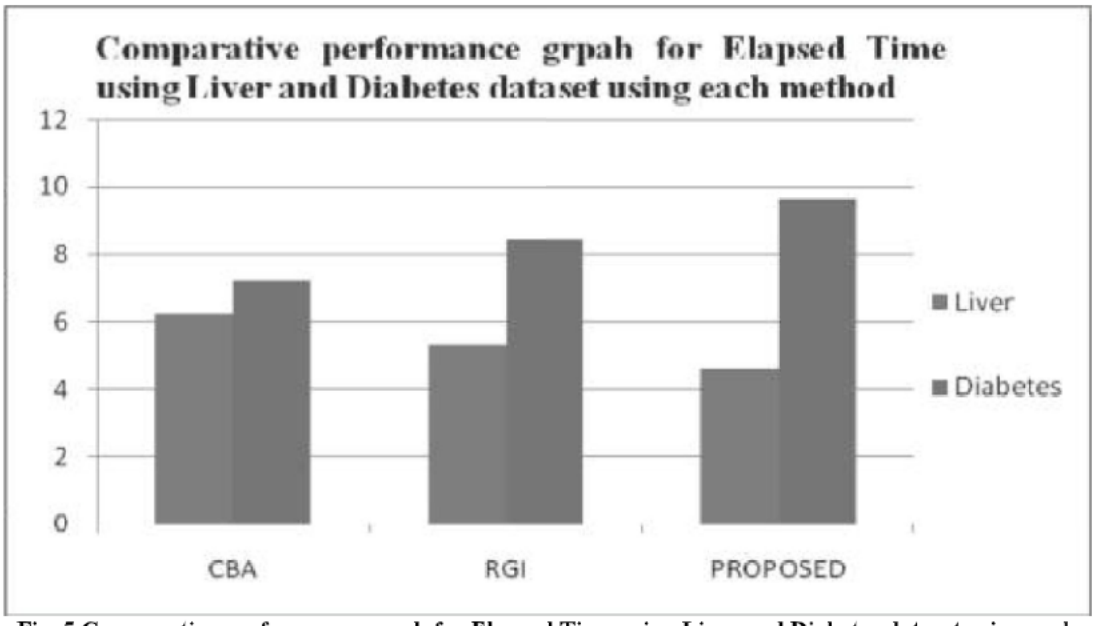

Fig. 5 Comparative performance graph for Elapsed Time using Liver and Diabetes dataset using each method

The Figure 2: shows comparative results of Liver data set, using CBA, RGI and Proposed method and here our proposed algorithm gives better result in the form of higher Accuracy and low Elapsed time as compared to the existing method.

Figure 3: shows that comparative result of Diabetes data set, using CBA, RGI using proposed method. Here our proposed algorithm shows better results in the form of higher Accuracy and low Elapsed time than the existing method.

Figure 4: shows comparative result of Accuracy for Liver and Diabetes data set, using CBA, RGI and Proposed method. Here our proposed algorithm shows better results in the form of higher Accuracy.

Figure 5: shows comparative result of Elapsed Time for Liver and Diabetes data set, using CBA, RGI and proposed method. Here our proposed algorithm shows better results in the form of low Elapsed time as compared to existing method.

V CONCLUSION

The Accuracy and Elapsed time was compared using proposed techniques against CBA & RGI for the same and different datasets. The analysis shows that results of Liver data set and Diabetes data set, using proposed method and proposed algorithm as compared to CBA & RGI give better result in the form of higher Accuracy and low Elapsed time than the existing method.

REFERENCES

[1] D. P. Shukla, Shamsher Bahadur Patel, Ashish Kumar Sen "A Literature Review in Health Informatics Using Data Mining Techniques", international journal of software & hardware research in engineering 2014 PP 123-129.

[2] Mihaela Gheorghe, Ruxandra Petre "Integrating Data Mining Techniques into Telemedicine Systems" Informatica Economică 2014 PP 120- 130.

[3] Li Jiang, Stefan M Edwards, Bo Thomsen, Christopher T Workman, Bernt Guldbrandtsen , Peter Sørensen "A random set scoring model for prioritization of disease candidate genes using protein complexes and data-mining of GeneRIF, OMIM and PubMed records" BMC Bioinformatics 2014 PP 1-13.

[4] Saurabh Pal, Vikas Chaurasia "Data Mining Approach to Detect Heart Diseases" IJACSIT 2013 PP 56-66

[5] Salim Diwani ,Suzan Mishol , Daniel S.Kayange ,Dina Machuve ,Anael Sam "Overview Applications of Data Mining In Health Care: The Case Study of Arusha Region" International Journal of Computational Engineering Research 2013 PP 73-77.

[6] Divya Tomar , Sonali Agarwal "A survey on Data Mining approaches for Healthcare" International Journal of Bio-Science and Bio-Technology 2013 PP 241-266.

[7] V.Krishnaiah, Dr.G.Narsimha, Dr.N.Subhash Chandra "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques" (IJCSIT) 2013, PP 39 – 45.

[8] Ashish Kumar Sen, Shamsher Bahadur Patel, Dr. D. P. Shukla "A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level" International Journal Of Engineering And Computer Science 2013 PP. 2663-2671.

[9] Shweta Kharya "Using Data Mining Techniques For Diagnosis And Prognosis Of Cancer Disease" International Journal of Computer Science, Engineering and Information Technology , 2012 PP 55-66

[10] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg "Data Mining in Clinical Decision Support Systems for Diagnosis, Prediction and Treatment of Heart Disease" International Journal of Advanced Research in Computer Engineering & Technology 2013 PP 218-223.

[11] Duen-Yian Yeh , Ching-Hsue Cheng, Yen-Wen Chen "A predictive model for cerebrovascular disease using data mining" Elsevier2011 PP 8970–8977

[12] M. Durairaj, V. Ranjani "Data Mining Applications In Healthcare Sector: A Study" International Journal Of Scientific & Technology Research 2013 PP 29-35.

[13] Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte "A Data Mining Approach For Prediction Of Heart Disease Using Neural Networks" IAEME 2012, PP. 30-40

[14] Menaouer Brahami, Baghdad Atmani ,Nada Matta "Dynamic knowledge mapping guided by data mining: Application on Healthcare" J Inf Process Syst, 2013 PP 1-30

[15] K. Rama Lakshmi , S.Prem Kumar "Utilization of Data Mining Techniques for Prediction and Diagnosis of Major Life Threatening Diseases Survivability-Review" International Journal of Scientific & Engineering Research, 2013 PP 923-932

[16] Mohammed Abdul Khaleel, Sateesh Kumar Pradhan, G.N.Dash "Finding Locally Frequent Diseases Using Modified Apriori Algorithm" International Journal of Advanced Research in Computer and Communication Engineering 2013 PP 3792-3797.

[17] Rui Zhang, Serguei Pakhomov, Janet T. Lee, Genevieve B. Melton "Navigating Longitudinal Clinical Notes with an Automated Method for Detecting New Information" IMIA and IOS Press 2013 PP 754-758.

[18] Mustafa Sofean ,Matthew Smith "A Real-Time Disease Surveillance Architecture Using Social Networks" MIE 2012.

[19] V. Krishnaiah, G. Narsimha, N. Subhash Chandra " A Study On Clinical Prediction Using Data Mining Techniques" International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) 2013, PP 239-248.

[20] Wu-Chen Su "A Preliminary Survey of Knowledge Discovery on Smartphone Applications (apps): Principles, Techniques and Research Directions for E-health" ICME 2014 PP 369-374.