Predicting heart failure using data mining with Rough set theory and
Fuzzy Petri Net

S. Meher Taj1, M. Sudha2, A. Kumaravel 3
1Assistant Professor, 2Research Scholar, Department of Mathematics
AMET Deemed to be University, Kanathur, Chennai-600112, India
3Dean, School of Computing
Bharath University, Selaiyur, Chennai-600073, India.

Abstract. The Rough Set Theory (RST) is a method that has proven its efficiency and simplicity in machine learning and successfully developing now a day’s vastly and rapidly. Fuzzy Petri nets (FPNs) are a potential modelling technique, which is used for knowledge representation and reasoning of rule-based expert systems. Though, RST has the efficiency in dimension reduction, it has to be proved with evidence by creating model using FPN with rule based reasoning. In this paper the induction of decision rules by RST executed with Fuzzy Petri Nets (FPN) is analyzed in the sense how it performed better than other data mining classifiers. The rule-based classifiers like jrip, part R, zero R are used for the comparison purposes. The knowledge captured from the rules through the best reduct has performed with the efficiency of RST and formulating rules by FPN. This paper experiments the Heart failure data to investigate the decision making from the rules generated by LERS system of RST with the approach of FPN. The heart failures during the follow up period of the patients are predicted with the pattern recognized form the data using the above process and the best evaluators are found during the experiments were pictured.

Keywords: Fuzzy Petri Nets, Rough set theory, Heart failure data, Data mining, Rule extraction, Prediction

1. Introduction

Machine learning is a mathematical technique that gives the machine the ability to construct a model for learning and enhancing the performance of a specific data function. This role hits the issues around supervised learning, clustering, reduction in dimensions and future prediction. Machine learning algorithms resolve the issues, and make decisions and predict results. Classification and regression are the function of supervised learning where clustering is not supervised [1]. Classification is the problem of defining a new pattern based on training data analysis and the pattern applied by the classification known as classifier. This role is widely needed because in the reduced space the data analysis such as classification and regression is more reliable and also the size of the data is large and becoming complicated in this advanced world. It also improves classifier performance, reduces time and space, and allows easy visualization in 2D or 3D. Techniques such as PCA, LDA, GDA, CCA etc. are used to solve this problem. The techniques like PCA, LDA, GDA, CCA etc., are been used to resolve this so far [2].

The Rough set theory has been noticed as an efficient tool in machine learning for the past decades and it has been successfully overcome the above underlined tasks in many fields [3, 4]. The RST-based algorithm called the LERS method being proof for it was announced by NASA's Johnson Space Center as the successful application of the RST in data mining by adopting LERS [8] as an expert system creation tool [5]. Some institutions have made use of this helpful software about the rough sets.
ROSE, ROSETTA, RSES, Rstudio [9] are the tools having the concepts of the Rough set theory and the algorithms based on that [10]. Among them ROSE tool has all the fundamentals of the Rough set theory. However, issues like too much running time, lack of current RST based algorithms, and low acceptance of big data are being found make the researchers felt lacking sometimes. In this paper, we try to figure out some issues faced by the beginners with the mentioned tool and with the concept of the Rough set theory

2. Methods and Materials

We apply several machine learning classifiers to predict the patient’s survival. RST can be described by means of lower and upper approximations. The set of granules is the universal set \( U \) and now the set is defined with respect to \( R \) where \( R \) is the equivalence relation assumed based on the knowledge prescribed in granules of \( U \). To describe the vague part of the set \( X \) with respect to \( R \), we need the approximations of rough set theory [11, 12].

2.1 Rule induction

LERS, the Rough Set Learning Examples is the successful minimal algorithm used to generate decision rules in the structure of if and then rules with a perfect pattern. Pattern is the knowledge calculated by all instances regarding the set of attributes and it can identify or test any instance that belongs to that knowledge. In three approaches, it induces rules; minimum set, exhaustive set, and satisfactory set. Maximum set produces a maximum number of rules that are adequate to convey all instances. Exhaustive set uses examples to induce all rules. Satisfactory setting leads to rules that satisfy user-defined requirements. In rule induction, it is important to process numerical attribute transformation into symbolic attribute. This process of discretization may vary by algorithm. The principles of decision and their algorithms have been discussed in [13]. Decision class approximations are described as three types, i.e. minimum set, exhaustive set and satisfactory set of decision rules [14]. For this study, ROSE2 [15] is a modular software system that uses the basic concepts of the rough set theory and the methods of rule finding.

2.2 Fuzzy Petri Nets

Petri Nets (PN) is a multi-system graphical and mathematical representation method. Promising tools are available to define and discover information science processes that are described as coincidental, asynchronous, distributed, parallel, non-deterministic and/or random [16,17]. Fuzzy Petri networks containing 2 kinds of nodes: Places and Transitions, where circles represent places and parallelogram represent transitions. Each place stands for an associate degree precedent or resultant and may or may not contain a token related with a degree of truth between zero and one that speaks the validity of the precedent or sequence for the amount of trust at intervals. Every transition that represents a rule is said to have an issue value of certainty between 0 and 1. The factor of certainty in rule represents the strength of the idea [18, 19].

2.3 Dataset

We analyzed a dataset containing the medical records of 299 heart failure patients collected from UCI repository [7,8]. The patients consisted of 105 women and 194 men, and their ages range between 40 and 95 years old. The details of the attributes are described below.
The above data was carried out for classification based on RST. For that we used the ROSE 2.2 tool which is completely constructed using the fundamentals of RST. Initially the RST calculated the approximations of the dataset based on the number of equivalent sets. The lower and upper approximation gives the quality of classification of the dataset which should not be changed or lessened while mining. Here the quality of classification of our dataset is shown below Fig 1.

Based on the RST reduction concept, here the CORE attributes are none because of vagueness not found in the boundary region. Also reducts are found using discernibility matrix which are detailed in the table. To find the best reduct we need to use frequency of attributes to get the equal number of atoms that are in the original set. Otherwise the reduct set will loss the original information.

The percentage of accuracy of the classification by stratified cross validation using basic minimal covering is 73.59%, which is found by the number of appropriately classified elements over the total amount of elements [1]. The number of correctly classified instances is 178 and incorrectly classified is 42. The other parameters like 50% similarity of partially matched rules with 21% of majority threshold in voting and rule strength in class support have given similar percentages of accuracy only vary in decimals. Among we found, the mentioned percentage is the highest.

Though the rough set is best for reduction, since it is a crisp set. The original scores well and need not to be reduced. But the frequency based reduct has found among the reducts created in the tool is shown in figure [3]. From that, we got five attributes are irrelevant (sex, anemia, diabetes, platelets and smoking) as they are very low in frequency of threshold 0.30. The accuracy of the same was calculated by the cross validation is not up to the original accuracy. It is decreased to 69.52%. so we

Table 1: Attributes description

| Feature                    | Explanation                                                                 | Measurement |
|----------------------------|----------------------------------------------------------------------------|-------------|
| Age                        | Age of the patient                                                         | Years       |
| Anaemia                    | Decrease of red blood cells or hemoglobin                                  | Boolean     |
| High blood pressure        | If a patient has hypertension                                              | Boolean     |
| Creatinine phosphokinase (CPK) | Level of the CPK enzyme in the blood                                      | mcg/L       |
| Diabetes                   | If the patient has diabetes                                                | Boolean     |
| Ejection fraction          | Percentage of blood leaving the heart at each contraction                  | Percentage  |
| Sex                        | Woman or man                                                              | Binary      |
| Platelets                  | Platelets in the blood                                                     | kiloplatelets/mL |
| Serum creatinine           | Level of creatinine in the blood                                           | mg/dL       |
| Serum sodium               | Level of sodium in the blood                                               | mEq/L       |
| Smoking                    | If the patient smokes                                                      | Boolean     |
| Time                       | Follow-up period                                                          | Days        |
| Class: death event         | If the patient died during the follow-up period                            | Boolean     |

The frequency based reduct found among the reducts created in the tool is shown in figure [3]. From that, we got five attributes are irrelevant (sex, anemia, diabetes, platelets and smoking) as they are very low in frequency of threshold 0.30. The accuracy of the same was calculated by the cross validation is not up to the original accuracy. It is decreased to 69.52%.
pushed to take the original set accuracy instead the reduct. Then using LERS, the rules generated are listed below. In that, we got 16 certain rules.

The above same Heart failure dataset was processing and classified using the data mining classifiers [21] like Jrip, PART, One R and Zero R for the comparison purposes using the data Mining tool WEKA 3.8 [22]. The classifiers output is described below in table 2.

Table 2: Classifier’s outputs for accuracy.

| Classifiers | Accuracy of classification | MAE  | Correctly classified objects | Incorrectly classified objects |
|-------------|----------------------------|------|------------------------------|-------------------------------|
| Jrip        | 71.91 %                    | 0.3823 | 215                          | 84                           |
| PART        | 67.89 %                    | 0.436  | 203                          | 96                           |
| Zero R      | 67.89 %                    | 0.436  | 203                          | 96                           |

The results of classifiers tabulated in the table 3 are the accuracy of them after the reducts. Here we have done the attribute selection evaluator using the ingrain which evaluates the worth of an attribute by measuring the information gain with respect to the class and used ranking search method. List of attributes with merit score after ranking is shown in figure 4. With threshold of 0.20, there are 7 attributes are removed out of 12. This ranking gives the reduct of 5 attributes (time, creatinine_phosphokinase, platelets, serum_creatinine, age). The same classifiers were carried out with this reduct and the results of the accuracy are tabulated in table 3. Except Jrip, other classifiers are giving the same percentage of accuracy. Whereas the Jrip is increased from 71.91% to 72.58% with less mean absolute error (MAE).The accuracy of rough set theory and other classifiers are compared in figure 5.

Table 3: Accuracy of selected attributes

| Classifiers | Accuracy of classification | MSE   |
|-------------|----------------------------|-------|
| Jrip        | 72.58 %                    | 0.3729 |
| PART        | 67.89 %                    | 0.436  |
| Zero R      | 67.89 %                    | 0.4365 |
2.4 Construction of Fuzzy Petri Net

It has been clearly seen that percentage of accuracy based on Rough set theory is better than the Jrip classifier. It has been done easily using RST without reduction. The 16 rules are generated using LERS system and listed below in the ‘If and then’ form.

R1: If (anaemia = 0) and (diabetes=1) and (high_blood_pressure =0 ) and (serum_creatinine =0) and ( sex=0) then DEATH EVENT =0
R2: If (Age =40,41,67,79,44,63,66,43,61,81,52,64,78,47,56) then DEATH EVENT =0
R3: If (anaemia = 0) and (diabetes=0) and (high_blood_pressure =0 ) and (serum_creatinine =135,137,139,140,133,144,141,130,128) then DEATH EVENT =0
R4: If (diabetes=0) and ( ejection_fraction=35,30,60,50) and (high_blood_pressure =1 ) and (smoking=0) then DEATH EVENT =0
R5: If ( time= 16, 22, 54, 71, 74, 75, 79, 80, 83, 85, 87, 91, 94, 104, 105, 106, 107, 108, 117, 118, 119, 120, 121, 145, 146, 147, 174, 175, 185, 186, 187, 188, 192, 197, 200, 201, 205, 208, 209, 210, 211, 212, 213, 215, 230, 233, 237, 240, 245, 246, 247, 250, 256, 270, 285) then DEATH EVENT =0
R6: If (diabetes =1) and (serum sodium= 142) and (smoking =1) then DEATH EVENT =0
R7: If (serum sodium=138, 137) and (smoking =0) then DEATH EVENT =0
R8: If (creatinine_phosphokinase= 291, 127, 224, 369, 737, 151, 211) then DEATH EVENT =0
R9: If (creatinine_phosphokinase=246, 7861, 146, 111, 160, 315, 123, 981, 168, 379, 149, 125, 128,220, 112, 70, 23, 249, 94, 855, 235, 124, 571, 588, 1380, 553, 577, 91, 3964, 260, 371, 789, 364, 7702, 318, 110, 161, 5882, 76, 280, 154, 328, 805, 943, 233, 2334, 2442, 776, 176, 395, 99, 145, 104, 1896, 418, 131, 427, 166, 2017, 258, 1199) then DEATH EVENT =1
R10: If (Age =45) and (diabetes=0) and (sex=1) and (smoking=0) then DEATH EVENT =1
R11: If (serum sodium =134) and (time=30, 172) then DEATH EVENT =1
R12: If (serum sodium=135) and (time=180) then DEATH EVENT =1

3. Discussion

The first 8 certain rules are bounded to the class 0 and remaining bounded to class 1. The decisions can be made between the rules since they were executed by FPN and the FPN chart represents that all
rules well fired without break seems the grouped rules and gives some précised decisions to predict the possibilities of Heart failure.

The set of rules represented a model by FPN, which can be used to characterize the uncertainty. The FPN model varies with expert knowledge beside with some parameters and more instances to create appropriate model to determine the risk factors about the instances. The predicted possibilities help to provide suitable treatment to the heart failure to avoid the risks of life.

The attributes “Time, age and creatinine phosphokinase” are the important factors found in this experiment for the heart failure. Both Jrip and the rough set theory have these three in the highest score which helps to conclude to get a pattern to predict the future cases.

In Figure 8, the Petri net template, transitions 1 to 16 reflect rules 1 to 16 in the implemented rule above, according to the quantities allocated to each place, and firing each transition ensures that the corresponding rule is fulfilled.

4. Conclusion

In our work, the analysis with the concept of Rough set theory has selected Time, age and creatinine phosphokinase as relevant attributes for this heart failure medical records which is confirmed by the relevance of RST and executed by the Fuzzy Petri net. Moreover our analysis shows that the approach of rough set theory can be used affectively in the health record of patients with cardiac diseases. Also the rules were execute in FPN conforms that the rules are precise and pruned. Since we found the pattern form the model after the decision rules of which have the best accuracy of prediction among all the experiments have done in this study, the mentioned attributes are the main factors for heart failure based on this record and can helps to predict the patients with the same issues to avoids the risks. As future development, this study has to be taken to the high dimensional data of other type of illness datasets.
Acknowledgement
The first author would like to thank the management of AMET Deemed to be University for their support and encouragement for this research study.

References:
[1]. Ian H.Witten and Eibe Frank, Data Mining Practical Machine Learning Tools and Techniques, Second Edition, Elsevier, pp 62-82.
[2]. C. Ding, X. He, H. Zha, H.D. Simon 2002 Adaptive Dimension Reduction for Clustering High Dimensional Data, Proceedings of International Conference on Data Mining.
[3]. Pawlak.Z, 2002 Rough Sets and Intelligent Data Analysis, Information Sciences, Vol147, no. 1, pp. 1-12.
[4]. Mienko R., Stefanowski, J., Taumi, K.&Vanderpooten, D.1996.Discovery-Oriented Induction of Decision Rules. Cahier du Lamsade, No.141, Université Paris Dauphine.
[5]. Grzymala-Busse, J.W 2004 Rough Set Theory with Applications to Data Mining, Real World Applications of Computational Intelligence. Springer, Heidelberg.
[6]. Ahmad T, Munir A, Bhatti SH, Aftab M, Raza MA,2017 Survival analysis of heart failure patients a case study. PLoS ONE12(7):0181001.
[7]. Ahmad T, Munir A, Bhatti SH, Aftab M, Ali Raza M,2019 Survival analysis of heart failure patients: a case study. Data set. Accessed. https://plos.figshare.com/articles/Survival_analysis_of_heart_failure_patients_A_case_study/5227684/1
[8]. Grzymala-Busse JW. LERS1992, A system for learning from examples based on rough sets. In: Slowinski R, editor. Intelligent decision support handbook of applications and advances of the rough sets theory. Kluwer Academic Publishers. pp 3-18.
[9]. Abbas, Z. and Burney, A.,2016 A Survey of Software Packages Used for Rough Set Analysis. Journal of Computer and Communications, 4, 10-18.
[10]. T. Slimani, 2013 Application of Rough Set Theory in Data Mining, International Journal of Computer Science & Network Solutions, vol. 1, no. 3, pp. 1-10.
[11]. Grzymala-Busse, J.W.,2004, Rough Set Theory with Applications to Data Mining, Real World Applications of Computational Intelligence. Springer, Heidelberg.
[12]. Zdzislaw Pawlak,2001.Rough Sets and Decision Algorithms, W. Ziarko and Y. Yao (Eds.): RSCTC 2000, LNAI 2005, pp. 30-45.
[13]. Polkowski L. & Skowron, A,1998 Rough Sets in Data Mining and Knowledge Discovery,Vol.1, PhysicaVerlag, pp 500-529.
[14]. ROSE Software http://idss.cs.put.poznan.pl/site/rose.html
[15]. Shen, V.R.L.,2006 Knowledge representation using high-level fuzzy Petri nets. IEEE T. Syst. Man Cy. A, 36(6): pp 1220-1227.
[16]. He, X., W.C. Chu, H. Yang and S.J.H. Yang.1999,A new approach to verify rule-based systems using petri nets. Proceeding of 23th IEEE Annual International Computer Software and Applications Conference (COMPSAC’99), pp: 462-467.
[17]. Jenitha.G and Kumaravel.A,2017 An Instance of Reinforcement Learning Based On Fuzzy Cognitive Maps, International Journal of Computational and Applied Mathematics. 12(1):200-207, 2017.
[18]. Paulraj Jayasimman I. Dhivya.T and Joseline Manora J 2018, Inverse Majority Neighborhood Number of a Graph, IOP Journal of Physics Conference Series .pp 1-6.
[19]. Jenitha.G,Paulraj Jayasimman I .Kumaravel.A ,2020 Energy value of Benzoxepine Derivatives International Journal of Scientific & Technology Research .pp:3137-3141.
[20]. Paulraj Jayasimman .I Jenitha .G, Kumaravel A ,2018 The Energy Graph for Minimum Majority Domination , International Journal of Engineering & Technology, pp: 562-565.