A GA approach for finding decision rules based on bireducts (Extended abstract)

Oleg Rybkin, Ivo Düntsch, Beatrice Ombuki-Berman

Brock University, St. Catharines, Ontario, L2S 3A1, Canada
kazkibergetic@gmail.com, duentsch@brocku.ca, bombuki@brocku.ca

1 Introduction

A decision system $\mathscr{D}$ in the sense of Pawlak [6] is a tuple $\langle U, A, d\rangle$, where U is a finite set of objects and A a finite set of attributes or features. Each $a \in A$ is considered as a function with domain U and some range V_a. The attribute $d \notin A$ is the decision attribute. A nonempty subset B of A is called *discerning* (with respect to d), if $b(x) = b(y)$ implies $d(x) = d(y)$ for all $x, y \in U, b \in B$. $\mathscr{D}$ is called *consistent* if it has a discerning set of attributes. A minimal discerning set is called a *decision reduct* or simply *reduct*. In traditional rough set theory, feature selection via reducts has been an important area. As finding reducts is computationally unfeasible [7], many methods of finding feature sets with an acceptable quality of classification have been proposed within classical rough set theory, see, for example, [1,5,8].

Recently, Ślęzak and Janusz [9] proposed to take into account not only the horizontal reduction of information by feature selection, but also a vertical reduction by considering suitable subsets of the original set of objects. Thus, the aim is to find areas in the two dimensional Object $\times$ Attribute plane which are in some sense best suited for classification. This leads to the following definition: A *(decision) bireduct* is a pair $\langle B, X\rangle$ such that $B \subseteq A, X \subseteq U$ and

R1. For all $b \in B, x, y \in X$, $b(x) = b(y)$ implies $d(x) = d(y)$; in this case, we write $B \Rrightarrow_X d$. (B is discerning all elements of X)
R2. If $C \subsetneq B$, there are $x, y \in X$ such that $c(x) = c(y)$ for all $c \in C$ and $d(x) \neq d(y)$. (Minimality of B with respect to X)
R3. If $X \subsetneq Y$, there are $x, y \in Y$ such that $b(x) = b(y)$ for all $b \in B$ and $d(x) \neq d(y)$. (Maximality of X with respect to B)

Decision rules now can be obtained as in classical rough set theory by restricting the scope of the quantifiers to the parts of a bireduct $\langle B, X\rangle$. Bireducts – which are consistent on their object set – may be viewed as inducing approximate rules on the whole object set U. As finding (optimal) bireducts is NP – hard, obtaining optimal solutions for this kind of problems is computationally intractable and thus heuristic methods are required for bireduct discovery in realistic time. In our present situation genetic algorithms (GAs) will be our method of choice for finding bidirects.

A GA is a search heuristic that imitates the process of natural selection [3,2]. Each of the chromosomes in the population is subjected to an evolutionary process until a suitable solution is found or the stopping condition is met. The chromosomes are then subjected to an iterative evolutionary process, that is, in each generation, fitness evaluation of each chromosome is done, and then the genetic operations, crossover, mutation and selection are applied on the chromosomes until the termination condition is met. In our work, tournament selection with elite retention is used to perform fitness-based selection. The order crossover (OX) [2] and the reciprocal exchange mutation genetic operations were used. A goal here is to minimize the number of attributes and maximize the number of objects that attributes are valid for, thus making it a multi-objective optimization problem and thus we proposed a multi-objective GA (MOGA) approach that uses Pareto Ranking [2] fitness evaluation strategy. By modifying the algorithm proposed

in [9], we got the algorithm for finding bireduct for provided GA chromosome. After our MOGA System generate all bireducts for the provided dataset we pass the bireducts to our Rough Sets System, that generates decision rules based on bireducts and applies it to the testing dataset.

2 Experimental discussion and conclusions

The Breast Cancer Wisconsin (Diagnostic) dataset from UCI Machine Learning Repository [4] which has 569 instances with 32 attributes was used. Cross-validation technique was employed where 512 instances were used for training and 57 instances for testing, using various empirically established GA parameters and performed 20 independent runs. To generate bireducts our MOGA system was used. The prediction accuracy for each run, as well as the number of bireducts used and total number of correct predictions was established whereby the average prediction accuracy over all 20 runs is 97%. When evaluating the the number of attributes and number of objects in each bireduct for each run received by using the MOGA system, it was shown that the system was able to significantly reduce the number of attributes without much reduction of the number of objects. A further analysis evaluated whether there is any dependency between the number of bireducts used in each run and the prediction accuracy. It was found that there is not significant difference in the number of bireducts used in each run and it does not affect the prediction quality.

In conclusion, following the work started by Ślęzak and Janusz [9], we proposed a new approach to generate bireducts using a multi–objective GA. The Pareto ranking scoring used in the GA is advantageous over a commonly used weighted approach for multi-objective optimization as it precludes the need for search for suitable weights a priori. Thus, compared to the research mentioned before, we do not need to provide the ratio value to the system, change it, and generate a huge amount of bireducts with different attributes/objects ratios. We were able to reduce the number of bireducts necessary for receiving a good prediction accuracy by using better quality bireducts provided by MOGA. Although the current results are encouraging, further analysis is underway using various data sets, as well as carrying out an empirical study comparing Pareto ranking fitness evaluation with other fitness evaluation techniques, in addition to incorporating various genetic operators.

References

1. Bazan, J.G., Nguyen, H.S., Nguyen, S.H., Synak, P., Wróblewski, J.: Rough set algorithms in classification problem. In: Polkowski, L., Tsumoto, S., Lin, T.Y. (eds.) Rough Set Methods and Applications, pp. 49–88. Physica Verlag, Heidelberg (2000)
2. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading, MA (1989)
3. Holland, J.H.: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence. MIT Press, Cambridge, MA, USA (1992)
4. Lichman, M.: UCI machine learning repository (2013), `http://archive.ics.uci.edu/ml`
5. Moshkov, M., Piliszczuk, M., B., Z.: Partial Covers, Reducts and Decision Rules in Rough Sets, Studies in Computational Intelligence, vol. 145. Springer Verlag, Heidelberg (2008)
6. Pawlak, Z.: Rough sets. International Journal of Computer & Information Science pp. 341–356 (1982)
7. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Słowiński, R. (ed.) Intelligent decision support: Handbook of applications and advances of rough set theory, System Theory, Knowledge Engineering and Problem Solving, vol. 11, pp. 331–362. Kluwer, Dordrecht (1992)
8. Ślęzak, D.: Rough sets and functional dependencies in data: Foundations of association reducts. In: Gavrilova, M., Tan, C., Wang, Y., Chan, K. (eds.) Transactions on Computational Science V, Lecture Notes in Computer Science, vol. 5540, pp. 182–205. Springer Verlag (2009), `http://dx.doi.org/10.1007/978-3-642-02097-1_10`
9. Ślęzak, D., Janusz, A.: Ensembles of bireducts: Towards robust classification and simple representation. In: Kim, T., Adeli, H. andŚlęzak, D., Sandnes, F., Song, X., Chung, K., Arnett, K. (eds.) Proceedings of FGIT 2011. Lecture Notes in Computer Science, vol. 7105, pp. 64–77. Springer Verlag (2011)