Journal of Engineering and Applied Sciences 15 (7): 1805-1808, 2020
ISSN: 1816-949X

An Intelligent Classifier for Group Decision Making Based on Rough Sets

[1]E.I. Elsedimy and [2]F. Algarni
[1]Department of Systems and Information Technology, Faculty of Management Technology and Information Technology, Port Said University, 42526 Port Said, Egypt
[2]Faculty of Computing and Information Technology, University of Bisha, 61922 Bisha, Saudi Arabia

Abstract: In this study, a new approach to combine multiple decision systems using multiple classifiers and rough sets methods is presented. This approach depends on our proposed algorithms that work to combine multiple decision systems by aggregating the lower and upper approximations. This improves the quality of decision rules by increasing the number of certain rules which enable us to make certain decisions. Our experiment results indicate that combining lower and upper approximations improves the quality of decision rules. Furthermore, it increases the classification accuracy computed by single and multiple classifiers compared to existing methods.

Key words: Rough set, decision making, multiple classifiers, quality of decision accuracy

INTRODUCTION

In last decade, several methods have been proposed for constructing a base classifier (Cruz *et al*., 2018; Nweke *et al*., 2019; Chowdhury *et al*., 2017). This field of research is known under several names: multiple classifiers, ensembles, consensus aggregation, committees and classifier fusion (Liang *et al*., 2014; Huang *et al*., 2012; Sun *et al*., 2008). Building a system of multiple classifiers is based on two steps. The first is how to create classifiers and the second is how to aggregate the prediction of the classifier's outputs. Several methods are used for creating multiple classifiers such as those which are introduced by Wozniak *et al*. (2014). All these strategies depend on the classification methods used by multiple classifiers and the aggregation methods for merging the classifier's outputs. The Bayesian approach is another aggregation method that selects the class with the highest posterior probability computed from training data. The classifier's outputs are aggregated using simple operators as minimum, maximum, average, product and ordered weighted averaging (Kuncheva *et al*., 2001; Liu *et al*., 2018). The predictions or the outputs of these multiple classifiers can be aggregated by several strategies, for example, the decision templates method. This method combines the classifier's outputs by comparing them to a characteristic template for each class. Fuzzy aggregation methods can be used for combining the classifier's outputs as the support for the class label. Simple operators such as minimum, maximum, average, product and weighted averaging are also used to aggregation the matrix of classifier's outputs while each item in the matrix represents the class label support.

The most important problem is how to deal with inconsistencies in decision rules (Alam *et al*., 2018; Cekik and Telceken, 2018; Zhan *et al*., 2017) where the object or feature may be classified by one decision rule as good and by another as medium. Traditional aggregation methods cannot deal with this problem. Thus, simple aggregation operators like minimum, maximum, average and product will fail to deal with this problem and it is not yet, sufficient to rely on. To address this problem, in this paper, rough set theory introduces a new method to replace the methods used to merge the classifier's prediction outputs by combining the lower and upper approximations to aggregate all the classifier's outputs and removing inconsistencies. The aim of this study is to improve the quality of decision rules derived from decision problems. We introduce an algorithm to combine the lower and upper approximation sets.

Rough set theory: Rough set theory was developed by Pawlak (1982, 1991) and Zhu (2009) in the early 1980s. It is a formal framework for the automated transformation of data into knowledge. In rough set theory, an Information System (IS) (Huang *et al*., 2012) is represented by a pair of the form IS = (U, R) where U is a non-empty finite set of objects called the universe and R is a non-empty finite set of attributes such that $r: U \rightarrow V_r$ for $r \in R$. V_r is called the domain of attribute a and defines the set of values that a can have. With any $P \subset R$, there is a P-indiscernibility relation denoted by IND(P) and can be defined as follows:

Corresponding Author: E.I. Elsedimy, Department of Systems and Information Technology, Faculty of Management Technology and Information Technology, Port Said University, 42526 Port Said, Egypt

$$IND(P) = \{(x, y) \in U^2 : \forall r \in P,\ r(x) = r(y)\} \tag{1}$$

From this definition, we constructed the equivalence classes of a set and the first step in the classification with rough sets. In the same context, the lower and upper approximation of a given set X can be defined, respectively as follows:

$$\begin{aligned} \underline{P}(X) &= \{x:[x]_p \subseteq X\}, \\ \overline{P}(X) &= \{x:[x]_p \cap X \neq \varnothing\} \end{aligned} \tag{2}$$

Decision tables: Decision tables (Zhan and Alcantud, 2019) are a special type of information system. More formally, if the attributes for each object are evaluated or measured by many users (experts), say n-users, we have n-decision tables $Dt_i = \{U, R \cup \{d_i\}, V, f_i\}$, i = 1, 2, ..., n. We can obtain n- indiscernability relations R_i, i = 1, 2, ..., n. Thus, we have n-lower approximations and n-upper approximations for a subset R of U, respectively as follows:

$$\text{N-lower approximations} = {}_n\underline{N}(R) = \bigcup_{i=1}^{n} \underline{p}_i(R) \tag{3}$$

$$\text{N-upper approximations} = {}^n\overline{N}(R) = \bigcap_{i=1}^{n} \overline{p}_i(R) \tag{4}$$

The accuracy measure of a set under n-decision systems can be characterized by the following coefficient:

$${}^{(n)}\alpha(R) = \frac{|{}_n\underline{N}(R)|}{|{}_n\overline{N}(R)|} \tag{5}$$

The proposed algorithm: The goal of the proposed algorithm is to enhance the decision-making capability of the knowledge generated by learning algorithms. It also aims to reduce data inconsistency by calculating the lower and upper approximations from large data sets. This algorithm induces rules from data based on rough set theory. For inconsistent data, a rough set induces two sets of rules: certain rules and possible rules. The final outputs are rules and their evaluation results. The basic idea for extracting decision rules from the original data based on a rough set depends on some reduction and rule induction algorithms. All of these algorithms extract decision rules from one decision system and make predictions in the form of "if then rules". These rules cannot cover all of the training examples. So, we introduce an approach using rough sets as a tool for combining multiple decision systems by aggregating the lower and upper approximations. Two types of rules can be induced from decision systems: certain and possible rules. Certain rules are generated from the lower approximation of unions of classes and possible rules are generated from the upper approximation of unions of classes. More formally, it can be defined as:

$$r_i : A_i \rightarrow y_i$$

where the left-hand side of the rule is called the rule antecedent which contains a conjunction of conditions attributes and the right-hand side is called the rule consequent which contains the predicate class y_i.

The proposed algorithm comprises three steps. In the first step, it splits the data set into sets of equal sizes. After this, it assigns each of the spilt data sets to one classifier that works based on rough set methodology. The second step is the core of the algorithm where we calculate the lower and upper approximation sets for each spilt data set and combine all the lower and upper approximation sets together in a new decision system. Finally, the last step starts by running test data to induce decision rules from the new decision system. The computation time of our proposed algorithm is $O(kn^2)$ for n-decision system.

The proposed algorithm:

Input : Decision table with multiple decision attributes ($a_1, a_2, \ldots, a_n$)
Output : Decision rules ($e_1, e_2, \ldots, e_n$)
//the training sets
1: For i = 1 to N do
2: Create the training set D_i by sampling U/N // N is the number of sampling
3: Train a base classifier C_i on D_i
//Determine equivalence classes based on condition attributes
4: Let $C = \{C_1^B, C_2^B, ..., C_n^B\}$ are the equivalence classes of the relation IND(B)
5: Set $\underline{N}(X) = \varnothing;\ \overline{N}(X) = \varnothing$
//Constructing equivalence classes based on decision attribute a
6: Set $X \backslash a_i = \{X_1^{ai}, X_2^{ai}, \ldots, X_n^{ai}\}$
//Computing lower and upper approximations sets
7: Set $\underline{L}(X) = \varnothing;\ \overline{U}(X) = \varnothing$
8: For j = 1 to n begin
9: if $X_j^{di} \subseteq X$ then $\underline{L}(X) = \underline{L}(X) \cup X_j^{di}$
10: else if $X_j^{di} \cap X = \varnothing$ then $\overline{U}(X) = \overline{U}(X) \cup X_j^{di}$
11: end if
//Combined lower and upper approximations sets
12: $\bigcup \underline{N}(X) = \underline{N}(X) \cup \underline{L}(X)$
13: $\bigcap \overline{N}(X) = \overline{N}(X) \cap \overline{U}(X)$
14: end for
15: end for
//Construct decision system
16: $DT = \bigcup \underline{N}(X) + \bigcap \overline{N}(X)$
17: Return decision rules for DT

MATERIALS AND METHODS

Experiments

Data and materials: We applied our approach to the data in Table 1 which was collected from different

Table 1: Experiment data sets

Data sets	No. examples	No. attributes	No. classes
Iris	150	5	3
Labor	57	17	2
Zoo	101	17	7
Weather	14	5	2
Soybean	683	36	19
Australian	690	14	2
Heart	270	14	2
Diabetes	768	9	2
Car	1728	7	4
Cancer	569	32	2

data sets at the University of California and others from data sets attached with WEKA Software. Our aim is to investigate the effectiveness of using rough set approximations on data set classification. Thus, we applied our proposed algorithm to calculate the classification accuracy for each data set and compare the results with our proposed algorithm.

RESULTS AND DISCUSSION

Our aim is to investigate the effectiveness of using rough set approximations on data set classification. Thus, we applied our proposed algorithm based on rough sets to calculate the lower and upper approximations sets. Two groups of experiments are conducted. In the first experiment, the classification accuracy for the proposed algorithm running on the data sets is measured.

The results are presented in Table 2 which shows that the lower approximation for the group of decision makers increases and the upper approximation decreases. This increase in the lower approximation leads to an increase in the number of certain rules which enables us to make certain decisions and improve the classification accuracy. On the other hand, our proposed algorithm has not derived any possible rules. The second observation is that the rough set classifier cannot handle continuous attributes and there is a need for a discretization method.

In the second experiment, we calculate the number of certain rules, possible decision rules and level of consistency. We verify whether our proposed algorithm achieves higher classification accuracy than individual classifiers. The rule quality measured after applying the proposed algorithm on the experiment data sets is depicted in Fig. 1 which shows that the number of certain rules increases and the number of possible rules decreases after the proposed algorithm is applied on all the data sets.

Table 2: The classification accuracy computed by the proposed algorithm

Decision makers	Lower approximation	Upper approximation	Classification accuracy (%)
Group 1	13	23	58.0
Group 2	11	25	40.33
Proposed algorithm	14	21	70.22

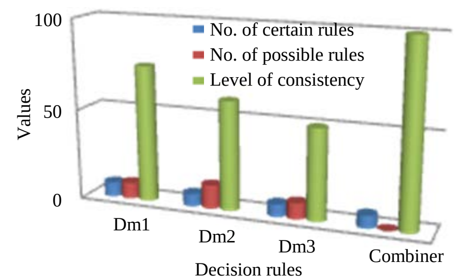

Fig. 1: The proposed algorithm improves the quality of the decision rules

CONCLUSION

This study presents an enhanced approach for group decision making based on rough set theory. Taking advantage of some useful proprieties of rough sets, we proposed an algorithm to combine the lower and upper approximations from group decision systems. The aim is to improve the quality of decision rules derived from decision problems. The proposed algorithm improves the quality of decision rules. Also, it increases the classification accuracy computed by single and multiple classifiers compared to existing methods.

REFERENCES

Alam, K.A., R. Ahmed, F.S. Butt, S.G. Kim and K.M. Ko, 2018. An uncertainty-aware integrated fuzzy AHP-WASPAS model to evaluate public cloud computing services. Procedia Comput. Sci., 130: 504-509.

Cekik, R. and S. Telceken, 2018. New Method Based on Rough Set for Filling Missing Value. In: Recent Developments and the New Direction in Soft-Computing Foundations and Applications, Zadeh, L., R. Yager, S. Shahbazova, M. Reformat and V. Kreinovich (Eds.). Springer, Cham, Switzerland, ISBN: 978-3-319-75407-9, pp: 41-48.

Chowdhury, A.K., D. Tjondronegoro, V. Chandran and S.G. Trost, 2017. Ensemble methods for classification of physical activities from wrist accelerometry. Med. Sci. Sports Exercise, 49: 1965-1973.

Cruz, R.M., R. Sabourin and G.D.C. Cavalcanti, 2018. Dynamic classifier selection: Recent advances and perspectives. Inf. Fusion, 41: 195-216.

Huang, J.T., M.H. Wang, W.J. Li and B. Gu, 2012. Multiple classifier fault diagnosis system based on dynamic weight. Acta Electronica Sin., 40: 734-738.

Kuncheva, L.I., J.C. Bezdek and R.P.W. Duin, 2001. Decision templates for multiple classifier fusion: An experimental comparison. Pattern Recognit., 34: 299-314.

Liang, S.Y., D.Q. Han and C.Z. Han, 2014. A novel diversity measure based on geometric relationship and its application to design of multiple classifier systems. Acta Autom. Sin., 40: 449-458.

Liu, Y., K. Qin and L. Martinez, 2018. Improving decision making approaches based on fuzzy soft sets and rough soft sets. Applied Soft Comput., 65: 320-332.

Nweke, H.F., Y.W. Teh, G. Mujtaba and M.A. Al-Garadi, 2019. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. Inf. Fusion, 46: 147-170.

Pawlak, Z., 1982. Rough sets. Int. J. Comput. Inf. Sci., 11: 341-356.

Pawlak, Z., 1991. Rough Sets: Theoretical Aspects of Reasoning about Data. 1st Edn., Kluwer Academic Publishers, London, UK., ISBN-13: 9780792314721.

Sun, L., C.Z. Han, J.J. Shen and N. Dai, 2008. Generalized rough set method for ensemble feature selection and multiple classifier fusion. Acta Autom. Sin., 34: 298-304.

Wozniak, M., M. Grana and E. Corchado, 2014. A survey of multiple classifier systems as hybrid systems. Inf. Fusion, 16: 3-17.

Zhan, J. and J.C.R. Alcantud, 2019. A novel type of soft rough covering and its application to multicriteria group decision making. Artif. Intell. Rev., 52: 2381-2410.

Zhan, J., Q. Liu and T. Herawan, 2017. A novel soft rough set: Soft rough hemirings and corresponding multicriteria group decision making. Applied Soft Comput., 54: 393-402.

Zhu, W., 2009. Relationship among basic concepts in covering-based rough sets. Inform. Sci., 179: 2478-2486.