Modified Soft Rough set for Multiclass Classification

S. Senthilkumar, H. Hannah Inbarani and S. Udhayakumar

Abstract Rough set theory has been applied to several domains because of its ability to handle imperfect knowledge. Most recent extension of rough set is soft rough set, where parameterized subsets of a universal set are basic building blocks for lower and upper approximations of a subset. In this paper, a new model of soft rough set, which is called modified soft rough set (MSR) where information granules are finer than soft rough sets, is applied for classification of medical data. In this paper, rough-set-based quick reduct approach is applied for selecting relevant features and MSR is applied for multiclass classification problem and the proposed work is compared with bijective soft set (BSS)-based classification, naïve Bayes, and decision table classifier algorithms based on evaluation metrics.

Keywords Soft rough set · Classification · Modified soft rough set · Quick reduct

1 Introduction

Classification and feature reduction are wide areas of research in data mining. Many practical applications of classification involve a large volume of data and/or a large number of features/attributes. Hence, it is necessary to remove irrelevant attributes and only the relevant attributes are used. The new idea is to solve multiclass classification problem with the modified soft rough set (MSR) method.

S. Senthilkumar (✉) · H. H. Inbarani · S. Udhayakumar
Department of Computer Science, Periyar University, Salem 636011, India
e-mail: pkssenthilmca@gmail.com

H. H. Inbarani
e-mail: hhinba@gmail.com

S. Udhayakumar
e-mail: uk2804@gmail.com

G. S. S. Krishnan et al. (eds.), *Computational Intelligence, Cyber Security and Computational Models*, Advances in Intelligent Systems and Computing 246,
DOI: 10.1007/978-81-322-1680-3_41,

In order to solve such a problem in medical diagnosis, attribute reduction, also called feature subset selection, is usually employed as a preprocessing step to select part of the attributes and focus the learning algorithm on relevant information. Rough set has strong ability in data processing and can extract useful rules from them [1]. The main aim of feature selection is to determine a minimal feature subset from a problem domain. A decision rule is a function, which maps an observation to an appropriate action. Deterministic rules correspond to the lower approximation, and non-deterministic rules correspond to the upper approximation [2].

Rough set theory and soft set theory are two different tools to deal with uncertainty. MSR sets satisfy all the basic properties of rough sets and soft sets. In some situations, equivalence relation cannot be defined, which is the basic requirement in rough set theory [3]. In these situations, MSR sets can help us to find approximations of subsets. Our proposed work consists of two parts: Initially, in the preprocessing stage, redundant data are removed and rules are derived from reduced data set. In this study, MSR-based classification is applied for generating decision rules from the reduced data set.

Modified Soft Rough Sets.

Soft set theory [4] deals with uncertainty and vagueness on the one hand, while on the other, it has enough parametrization tools. Feng et al. [5] introduced soft rough sets and comparative analysis of rough sets and soft sets. In order to strengthen the concept of soft rough sets, a new approach called MSR is presented in [6]. The definitions for lower and upper soft rough approximations are given below:

Definition 1 Let (F, A) be a soft set over U, where F is a map $F: A \rightarrow P(U)$. Let $u: U \rightarrow P(A)$ be another map, defined as $u(x) = \{a : x \in F(a)\}$. Then, the pair (U, u) is called MSR approximation space, and for any $X \subseteq U$. Ulower MSR approximation is defined as Table 1.

$$\underline{X}\phi = \{X \in U : \phi(x) \neq \phi(y), \quad \text{for all}\, y \in x^c \tag{1}$$

where $X^c = U - X$ and its upper MSR approximation is defined as follows:

$$\bar{X}\phi = \{X \in U : \phi(x) = \phi(y), \quad \text{for some}\, y \in x \tag{2}$$

If $\underline{X}\phi \neq \bar{X}\phi$, then X is said to be an MSR set.

2 Methodology

The methodology adopted in this work is given in Fig. 1. The MSR-based classification approach is applied for generating rules from the trained data, and rule matching is applied for test data to compute the decision class based on reliability analysis. In this study, the proposed approach is applied for medical diagnosis [7].

Table 1 Sample data for MSR sets

U	E1	E2	E3	D
S1	1	1	0	1
S2	0	1	0	0
S3	0	1	0	1
S4	1	0	0	0
S5	0	0	1	0
S6	1	0	1	1

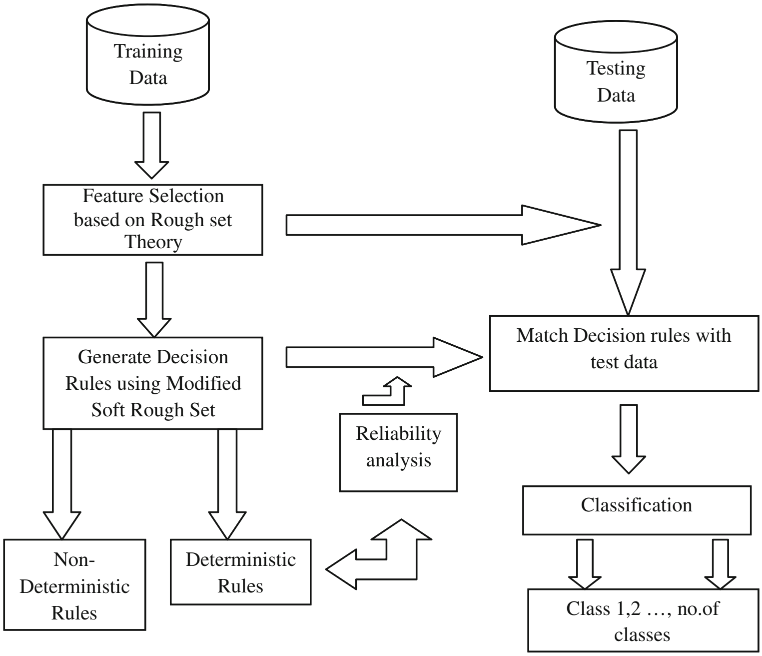

Fig. 1 Overall structure of the proposed work

3 MSR-Based Classification Algorithm

MSR-based classification algorithm is shown in Fig. 2. In this approach, lower and upper soft rough approximations of the given data set based on decision class X are constructed. In the second step, AND operation is applied to combine the soft sets. In the third step, deterministic rules are generated based on lower soft rough approximation. In the fourth step, non-deterministic rules are generated based on upper soft rough approximation and support for each non-deterministic rule is computed using step 5 of the algorithm.

The proposed algorithm is explained with example given in Table 1.

Step 1: Construct Modified Soft Rough Set (F, E) from the Dataset given in table 1.

Step 2: Apply AND operation for all conditional attributes (F, E1) ∧ (F, E2) ∧ (F, E3)ф(s1) = {e1,e2},ф (s2) = {e2} = ф (s3),ф (s4) = {e1}, ф (s5) ={e3},ф (s6) = {e1, e3}.

Step 3: Generate deterministic rules by using

$$\underline{X}\phi = \{X\epsilon U: \phi(x) \neq \phi(y), for\ all\ y\epsilon x^c$$

x={s1,s3,s6}, x^c={s2, s4,s5},$\underline{X}\phi$ ={s1,s6}

If E1= 1 and E2= 1and E3= 0 =>d= 1

If E1= 1 and E2= 0and E3= 1 =>d= 1

Step 4: Generate non-deterministic rules using

$$\overline{X}\phi = \{X\epsilon U: \phi(x) = \phi(y), for\ some\ y\epsilon x$$

x={s1,s3,s6}, x^c={s2,s4,s5}, $\overline{X}\phi$={s1,s2,s3,s6}

If E1= 1 and E2= 1and E3= 0 =>d= 1

If E1= 0 and E2= 1and E3= 0 =>d= 0

If E1= 0 and E2= 1and E3= 0 =>d= 1

If E1= 1 and E2= 0and E3= 1 =>d= 1

Step 5: Compute the support value for each non-deterministic rule.

$$\text{Support} = \left(\frac{1}{2} \wedge \frac{1}{2}\right) = 0.5$$

4 Experimental Analysis

Classification is a data mining technique used to predict group membership for data instances. The model is used to classify new objects [7]. In this work, classification accuracy of the proposed approach is compared with three different classifiers BSS classification, naïve Bayes, and decision table using the accuracy

Algorithm: MSR based Classification
Input: Given Dataset with conditional attributes 1, 2, … ,n-1 and the Decision attribute n.
Output: Generated Decision Rules for multiclass values.
Step1: Construct MSR approximation space for the given Dataset
Step2: Apply AND operation for all conditional attributes.
Step3: Generate deterministic rules using

$$\underline{X}\phi = \{X \in U: \phi(x) \neq \phi(y), for\ all\ y \epsilon x^c$$

Step4: Generate non-deterministic rules by using

$$\overline{X}\phi = \{X \epsilon U: \phi(x) = \phi(y), for\ some\ y \in x$$

Step5: Compute the support value for each non-deterministic rule

$support = \frac{support\ (A \wedge B)}{support\ (A)}$ Where A is the description on condition attributes and B the description on decision attributes.

Fig. 2 MSR-based classification algorithm

Table 2 Performance analysis of classification algorithms for breast cancer and hepatitis

	Breast cancer				Hepatitis			
	MSR sets	Bijective soft set	Decision table	Naïve Bayes	MSR sets	Bijective soft set	Decision table	Naïve Bayes
Precision	0.983	0.972	0.954	0.956	0.848	0.762	0.72	0.747
Recall	0.980	0.961	0.95	0.950	0.735	0.727	0.66	0.717
F-measure	0.981	0.96	0.951	0.953	0.731	0733	0.637	0.71

Table 3 Performance analysis of classification algorithms for Pima Indian diabetes and liver data sets

	Pima Indian diabetes				Liver			
	MSR sets	Bijective soft set	Decision table	Naïve Bayes	MSR sets	Bijective soft set	Decision table	Naïve Bayes
Precision	0.836	0.805	0.758	0.769	0.986	0.981	0.684	0.673
Recall	0.792	0.799	0.766	0.773	0.99	0.976	0.671	0.594
F-measure	0.802	0.796	0.758	0.771	0.988	0.972	0.678	0.59

measures precision, recall, and F-measure. The four medical data sets taken for the experimental analysis of the proposed approach are taken from UCI repository (www.ics.uci.edu/ ~ mlearn/).

From the results provided in Tables 2 and 3, it can be easily concluded that the MSR algorithm is an effective method for medical diagnosis. The proposed classification method is applied to reduced data set, thus reducing the number of rules, which leads to significantly improved classification accuracy and results in a significant difference in a patient's chance for recovery.

5 Conclusion

In this paper, MSR set approach is proposed for engendering deterministic and non-deterministic rules for the multiclass classification of medical data set. Comparison of the proposed approach with familiar BSS classification, decision table, and naïve Bayes classifier algorithms is carried out using performance metrics. The better results show that MSR-based classification is suitable for medical data. Therefore, future treatment decision based on the MSR sets of a medical data is a rational approach toward preventing the outgrowth of metastases.

References

1. C. Velayutham and K. Thangavel, "Unsupervised Quick Reduct Algorithm Using Rough Set Theory", International Journal of electronic science and technology, Vol. 9, pp. 193-201, 2011.
2. L. P. Khoo, S. B. Tor and L. Y. Zhai, "A Rough-Set-Based Approach for Classification and Rule Induction, "International Journal of Advanced Technology, Vol. 15, pp. 438–444, 1999.
3. Z. Pawlak, "Rough sets", International Journal of Computational Information Science, Vol. 11, pp. 341–356, 1982.
4. Molodtsov, "Soft set theory-RST results", Computational Mathematical Application. Vol.37, pp. 19-31, 1999.
5. FengFeng, Xiaoyan Liu, VioletaLeoreanu-Fotea, Young Bae Jun "Soft sets and soft rough sets", Information Sciences, Vol. 181, pp. 1125–1137, 2011.
6. Muhammad Shabir, Muhammad Irfan Ali, TanzeelaShaheen, "Another approach to soft rough sets", Knowledge-Based Systems, Vol. 40, pp. 72–80, 2013.
7. S.Udhaya Kumar, H. Hannah Inbarani, S.Senthilkumar, "Bijective soft set based classification of Medical data", International Conference on PRIME, pp. 517 – 521. 2013.