An Efficient Fuzzy-Rough Attribute Reduction Approach

Yuhua Qian, Chao Li, and Jiye Liang

Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, 030006, Shanxi, China
jinchengqyh@126.com, {lichao,ljy}@sxu.edu.cn

Abstract. Fuzzy rough set method provides an effective approach to data mining and knowledge discovery from hybrid data including categorical values and numerical values. However, its time-consumption is very intolerable to analyze data sets with large scale and high dimensionality. In this paper, we propose a strategy to improve a heuristic process of fuzzy-rough feature selection. Experiments show that this modified algorithm is much faster than its original version. It is worth noting that the performance of the modified algorithm becomes more visible when dealing with larger data sets.

Keywords: fuzzy rough sets, attribute reduction, algorithm.

1 Introduction

Feature selection, also called attribute reduction, is a common problem in pattern recognition, data mining and machine learning. Unlike other dimensionality-reduction methods, attribute reduction preserves the original meaning of the features after reduction.

In order to deal with numerical and categorical data (or a mixture of both) in data sets, fuzzy rough set model was first proposed by Dubois and Prade [1], which combines rough set and fuzzy set together. To widely apply the method, many extended versions and relative applications have been developed, cf. [5], [13], [14], [16], [17], [19], [21], [22]. The lower/upper approximation in these fuzzy rough set models tries to give a membership function of each object to a set.

Attribute reduction using fuzzy rough sets is often called fuzzy-rough feature selection. To support efficient feature selection, many heuristic algorithms have been developed in fuzzy rough set theory, cf. [2], [4]-[6], [8]-[10], [22]. Each of these feature selection methods can extract a single reduct from a given decision table. For convenience, from the viewpoint of heuristic functions, we classify these feature selection methods into two categories: fuzzy positive-region reduction and fuzzy information entropy reduction. In this paper, we only study fuzzy positive-region reduction.

The concept of positive-region was proposed by Pawlak in [15], which is used to measure the significance of a condition attribute in a decision table. Then, Hu and Cercone [7] proposed a heuristic attribute reduction method, called

J.T. Yao et al. (Eds): RSKT 2011, LNCS 6954, pp. 63–70, 2011.

positive-region reduction, which remains the positive region of target decision unchanged. Under Dubois's fuzzy rough set model, Jensen and Shen [8]-[10] developed a series of heuristic fuzzy-rough feature selection algorithms based on fuzzy positive-region. Bhatt and Gopal [2] proposed a modified version to improve computational efficiency. Under Hu's fuzzy rough set model and Wang's fuzzy rough model, Hu et al. [5] extended the method from the literature [7] to select a feature subset from hybrid data. Owing to the consistency of ideas and strategies of these methods, we regard the method from [5] as their representative.

In a recent published paper in Artificial Intelligence, to overcome the shortcoming of computationally time-consuming of all heuristic attribute reduction algorithms, Qian et al. [18] proposed an accelerator for attribute reduction in rough set theory, which is based on a theoretical framework called positive approximation. Using the experience of the method for reference, in this paper, we wish to develop an extended version of the accelerator for accelerating fuzzy-rough feature selection. By incorporating the new accelerator into fuzzy-rough heuristic attribute reduction method, we construct its modified version.

The study is organized as follows. Fuzzy rough set models are briefly reviewed in Section 2. In Section 3, we establish the forward approximation framework, and develop a modified attribute reduction algorithm based on the forward approximation. In Section 4, experiments on three public data sets show that this modified algorithm is much faster original version. Finally, Section 5 concludes this paper.

2 Preliminaries

In this section, we review several fuzzy-rough set models and some related concepts.

Given a nonempty finite set U, $\widetilde{R}$ is a fuzzy binary relation over U, the granular structure of the universe generated by a fuzzy binary relation $\widetilde{R}$ is defined as

$$\langle U, \widetilde{R} \rangle = ([x_1]_{\widetilde{R}}, [x_2]_{\widetilde{R}}, \cdots, [x_n]_{\widetilde{R}}), \tag{1}$$

where $[x_i]_{\widetilde{R}} = r_{i1}/x_1 + r_{i2}/x_2 + \cdots + r_{in}/x_n$. $[x_i]_{\widetilde{R}}$ is the fuzzy neighborhood of x_i and r_{ij} is the degree of x_i equivalent to x_j. Here, "+" means the union of elements. The cardinality of $[x_i]_{\widetilde{R}}$ can be calculated with $|[x_i]_{\widetilde{R}}| = \sum_{j=1}^{n} r_{ij}$, which appears to be a natural generalization of the cardinality of a crisp set.

In this case, $[x_i]_{\widetilde{R}}$ is a fuzzy set and the family of $[x_i]_{\widetilde{R}}$ forms a fuzzy concept system of the universe. This system will be used to approximate the object subset of the universe.

Let $\widetilde{X}$ be a fuzzy set. Then, it can be represented as

$$\widetilde{X} = \mu_{\widetilde{X}}(x_1)/x_1 + \mu_{\widetilde{X}}(x_2)/x_2 + \cdots + \mu_{\widetilde{X}}(x_n)/x_n, \tag{2}$$

where $\mu_{\widetilde{X}}(x_j)$ denotes the membership degree of the object x_j in $\widetilde{X}$.

The first fuzzy rough set model was introduced by Dubois and Prade [4]. By their definition, a universe of objects $U = \{x_1, x_2, \ldots, x_n\}$ is described by a fuzzy binary relation $\widetilde{R}$. Given $X \subseteq U$ a crisp subset of objects, the memberships of an object x_i in a fuzzy rough set $(\underline{\widetilde{R}}(X), \overline{\widetilde{R}}(X))$ of fuzzy sets on U are described as

$$\begin{cases} \mu_{\underline{\widetilde{R}}(X)}(x_i) = \inf_{x_j \in U} \max\{1 - \widetilde{R}(x_i, x_j), \mu_X(x_j)\}, \\ \mu_{\overline{\widetilde{R}}(X)}(x_i) = \sup_{x_j \in U} \min\{\widetilde{R}(x_i, x_j), \mu_X(x_j)\}, \end{cases}$$

where U is a nonempty universe and R is a fuzzy binary relation on U.

In a recent paper, Hu et al. [9] gave another definition of a fuzzy rough set in the context of hybrid data (for simplification, called Hu's fuzzy rough set), which is shown as follows.

Let $\langle U, \widetilde{R} \rangle$ be a fuzzy approximation space and $\widetilde{X}$ a fuzzy subset of U. The lower approximation $\underline{\widetilde{R}}\widetilde{X}$ and upper approximation $\overline{\widetilde{R}}\widetilde{X}$ are defined as [9]

$$\begin{cases} \underline{\widetilde{R}}\widetilde{X} = \{x_i \mid [x_i]_{\widetilde{R}} \subseteq \widetilde{X},\ x_i \in U\}, \\ \overline{\widetilde{R}}\widetilde{X} = \{x_i \mid [x_i]_{\widetilde{R}} \cap \widetilde{X} \neq \emptyset,\ x_i \in U\}, \end{cases}$$

where $[x_i]_{\widetilde{R}} \subseteq \widetilde{X}$ means $\mu_{[x_i]_{\widetilde{R}}}(x_i) \leq \mu_{\widetilde{X}}(x_i)$, and $[x_i]_{\widetilde{R}} \cap \widetilde{X} \neq \emptyset$ implies that $\min\{\mu_{[x_i]_{\widetilde{R}}}(x_i), ; \mu_{\widetilde{X}}(x_i)\} \neq 0$, $\emptyset = \{\frac{0}{x_1} + \frac{0}{x_2} + \cdots + \frac{0}{x_n}\}$. The order pair $\langle \underline{\widetilde{R}}\widetilde{X}, \overline{\widetilde{R}}\widetilde{X} \rangle$ is called a fuzzy rough set. Without loss of generality, we will select Hu's fuzzy rough model as their representative in this study.

A decision table is an information system $S = (U, C \cup D)$, where C is called a condition attribute set and D is called a decision attribute [8]. In practical decision-making issues, in general, the decision attribute D can induce an equivalence partition, i.e., a crisp classification. In this paper, we only focus on this kind of decision tables. Assume the objects are partitioned into r mutually exclusive crisp subsets $\{Y_1, Y_2, \cdots, Y_r\}$ by the decision attribute D. Given any subset $B \subseteq C$ and $\widetilde{R}_B$ is the fuzzy similarity relation induced by B, then one can define the lower and upper approximations of the decision attribute D as

$$\begin{cases} \underline{\widetilde{R}_B}D = \{\underline{\widetilde{R}_B}Y_1, \underline{\widetilde{R}_B}Y_2, \cdots, \underline{\widetilde{R}_B}Y_r\}, \\ \overline{\widetilde{R}_B}D = \{\overline{\widetilde{R}_B}Y_1, \overline{\widetilde{R}_B}Y_2, \cdots, \overline{\widetilde{R}_B}Y_r\}. \end{cases}$$

Denoted by $POS_B(D) = \bigcup_{i=1}^{r} \underline{\widetilde{R}_B}Y_i$, it is called the positive region of D with respect to the condition attribute set B.

3 An Efficient Algorithm to Fuzzy-Rough Feature Selection

Definition 1. *Let $P = \{\widetilde{R}_1, \widetilde{R}_2, \cdots, \widetilde{R}_n\}$ be a family of fuzzy binary relations with $\widetilde{R}_1 \succeq \widetilde{R}_2 \succeq \cdots \succeq \widetilde{R}_n$, and X a crisp set. Given $P_i = \{\widetilde{R}_1, \widetilde{R}_2, \cdots, \widetilde{R}_i\}$, we define P_i-lower approximation $\underline{P_i}(X)$ and P_i-upper approximation $\overline{P_i}(X)$ of P_i-positive approximation of X as*

$$\begin{cases} \underline{P_i}(X) = \bigcup_{k=1}^{i} \underline{\widetilde{R}_k}X_k, \\ \overline{P_i}(X) = \overline{\widetilde{R}_i}X, \end{cases}$$

where $X_1 = X$ and $X_k = X - \bigcup_{j=1}^{k-1} \underline{\widetilde{R}_j}X_j$, $k = 2, 3, \cdots, n$, $i = 1, 2, \cdots, n$.

Definition 2. *Let $P = \{\widetilde{R}_1, \widetilde{R}_2, \cdots, \widetilde{R}_n\}$ be a family of fuzzy binary relations with $\widetilde{R}_1 \succeq \widetilde{R}_2 \succeq \cdots \succeq \widetilde{R}_n$ and $U/D = \{Y_1, Y_2, \cdots, Y_r\}$. Lower approximation and upper approximation of D with respect to P_i are defined as*
$\begin{cases} \underline{P_i}D = \{\underline{P_i}(Y_1), \underline{P_i}(Y_2), \cdots, \underline{P_i}(Y_r)\}, \\ \overline{P_i}D = \{\overline{P_i}(Y_1), \overline{P_i}(Y_2), \cdots, \overline{P_i}(Y_r)\}. \end{cases}$ *$\underline{P_i}D$ is also called the positive region of D with respect to the granulation order P_i, denoted by $POS^U_{P_i}(D) = \bigcup_{k=1}^{r} \underline{P_i}Y_k$.*

Theorem 1. ***(Recursive expression principle)*** *Let $P = \{\widetilde{R}_1, \widetilde{R}_2, \cdots, \widetilde{R}_n\}$ be a family of fuzzy binary relations with $\widetilde{R}_1 \succeq \widetilde{R}_2 \succeq \cdots \succeq \widetilde{R}_n$ and $U/D = \{Y_1, Y_2, \cdots, Y_r\}$. Given $P_i = \{\widetilde{R}_1, \widetilde{R}_2, \cdots, \widetilde{R}_i\}$, we have*
$$POS^U_{P_{i+1}}(D) = POS^U_{P_i}(D) \cup POS^{U_{i+1}}_{\widetilde{R}_{i+1}}(D),$$
where $U_1 = U$ and $U_{i+1} = U - POS^U_{P_i}(D)$.

As Shannon's information entropy [20] was introduced to search reducts in classical rough set model. In fact, several authors also have used variants of Shannon's entropy to measure uncertainty in rough set theory and construct heuristic algorithm of attribute reduction [1], [3], [11], [12], [19]. Hu, Xie and Yu [6] proposed a so-called fuzzy information entropy to fuzzy rough set model and used its fuzzy conditional entropy to design a heuristic feature selection algorithm. This reduction method remains the fuzzy conditional entropy of target decision unchanged, denoted by FSCE, in which the fuzzy conditional entropy reads as
$$H(D|B) = -\frac{1}{|U|} \sum_{i=1}^{|U|} \log \frac{|[x_i]_{\widetilde{R}_B} \cap [x_i]_{\widetilde{R}_D}|}{|[x_i]_{\widetilde{R}_B}|}.$$
Using the fuzzy conditional entropy, the definitions of the significance measures are expressed in the following way.

Definition 3. *Let $S = (U, C \cup D)$ be a decision table, $B \subseteq C$ and $\forall a \in C - B$. The significance measure of a in B is defined as*
$$Sig(a, B, D, U) = H(D|B) - H(D|B \cup \{a\}).$$

In a heuristic fuzzy-rough feature selection algorithm, based on the above theorems, one can find an attribute reduct by gradually adding selected attributes.

From the discussion in the previous subsection, we can construct an improved forward search algorithm based on the forward approximation, which is formulated as follows.

Algorithm 1. A general improved feature selection algorithm based on the forward approximation (FA)

Input: Decision table $S = (U, C \cup D)$;

Output: One reduct *red*.

Step 1: $red \leftarrow \emptyset$, $i \leftarrow 1$, $R_1 \leftarrow red$, $P_1 \leftarrow \{R_1\}$ and $U_1 \leftarrow U$; //*red* is the pool to conserve the selected attributes

Step 2: While $EF(red, D) \neq EF(C, D)$ Do //This provides a stopping criterion.

{ Compute the positive region of positive approximation $POS^U_{P_i}(D)$,

$U_{i+1} \leftarrow U - POS_{P_i}^{U}(D)$,
$i \leftarrow i+1$,
$B \leftarrow C - red$,
Select $a_0 \in B$ which satisfies $Sig(a_0, red, D, U_i) = max\{Sig(a_k, red, D, U_i), a_k \in B\}$,
If $Sig(a_0, red, D, U_i) > 0$, then $red \leftarrow red \cup \{a_0\}$,
$R_i \leftarrow R_i \cup \{a_0\}$,
$P_i \leftarrow \{R_1, R_2, \cdots, R_i\};\}$
Step 3: Return *red* and End.

Computing the significance measure of an attribute $Sig^{inner}(a_k, C, D, U)$ is one of the key steps in FA, which time complexity is $O(|C||U|^2)$. In Step 2, we begin with the empty set and add an attribute with the maximal significance into the set in each stage until finding a reduct. This process is called a forward reduction algorithm whose time complexity is $O(\sum_{i=1}^{|C|}(|C|-i+1)|U_i|^2)$. However, the time complexity of the original heuristic algorithm is $O(\sum_{i=1}^{|C|}(|C|-i+1)|U|^2)$. Obviously, the time complexity of FA is much lower than that of each of classical heuristic attribute reduction algorithms. Hence, one can draw a conclusion that the general feature selection algorithm based on the forward approximation (FA) may significantly reduce the computational time for fuzzy-rough feature selection.

4 Experimental Analysis

Some heuristic attribute reduction methods have been developed for hybrid data, cf. [2], [4]-[6], [8]-[10], [22]. The objective of the following experiment is to show the performance of time reduction of the improved algorithm. The data used in the experiments are outlined in Table 1, which were all downloaded from UCI Repository of machine learning databases.

Table 1. Data sets description

Data sets	Samples	Features	Classes
Image Segmentation	2310	19	7
Sonar, mines vs. rocks	208	60	2
Wisconsin diagnostic breast cancer	569	30	2

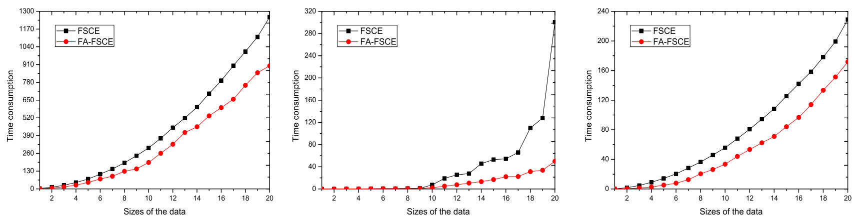
For numeric data, we normalize the numerical attribute a into the interval $[0,1]$ with $a' = \frac{a-a_{min}}{a_{max}-a_{min}}$. The value of the fuzzy similarity degree r_{ij} between objects x_i and x_j with respect to numerical attribute a is computed as

$$r_{ij} = \begin{cases} 1 - 4 \times |x_i - x_j|, |x_i - x_j| \leq 0.25, \\ 0, \qquad otherwise. \end{cases}$$

As $r_{ij} = r_{ji}$ and $r_{ii} = 1$, $0 \leq r_{ij} \leq 1$, the matrix $M = (r_{ij})_{n \times n}$ is a fuzzy similarity relation.

Table 2. The time and attribute reduction of the algorithms FSCE and FA-FSCE

Data sets	FSCE algorithm		FA-FSCE algorithm	
	Selected features	Time (s)	Selected features	Time (s)
Image Segmentation	17	1258.0468	17	900.5781
Sonar, mines vs. rocks	41	300.5625	41	50.0000
Cancer1	27	228.9218	27	171.8750



(a) (b) (c)

Fig. 1. Times of FSCE and FA-FSCE versus the size of data

From the definition of attribute reduction based on fuzzy rough sets, we know that the modified attribute reduction algorithm must select an attribute reduct from original attributes. Therefore, in the following experiment, we only consider attribute reducts obtained and computational time.

In what follows, we compare FSCE with FA-FSCE on those six real world data sets shown in Table 1 from computational time and selected feature subsets. Table 2 presents the comparisons of selected features and computational time with original algorithm FSCE and the accelerated algorithm FA-FSCE on three data sets. While Figure 1 gives more detailed change trendline of the algorithms with size of data set becoming increasing.

From Table 2 and Figure 1, it is easy to see that the modified algorithm is consistently faster than its original counterpart. Sometimes, the reduced time can almost achieves five-fifths of the original computational time. For example, the reduced time achieves 250.5625 seconds on the data set (Sonar, mines vs. rocks). Furthermore the differences are profoundly larger when the size of the data set increases. Hence, attribute reduction based on the accelerator should be a good solution.

5 Conclusions

In this study, a theoretic framework based on fuzzy rough set theory have been proposed, called the forward approximation, which can be used to accelerate algorithms of heuristic attribute reduction. Based on this framework, a general heuristic feature selection algorithm has been presented. A heuristic fuzzy-rough attribute reduction algorithm has been revised and modified. Experimental studies pertaining to three UCI data sets show that the modified algorithm can significantly reduce computing time of attribute reduction.

Acknowledgements. This work was supported by the national natural science foundation of China (No. 60903110, 60773133, No. 70971080), and the Natural science fund of Shanxi province, China (No. 2009021017-1 and 2008011038).

References

1. Beaubouef, T., Perty, F.E., Arora, G.: Information-theoretic measures of uncertainty for rough sets and rough relational databases. Inf. Sci. 109, 185–195 (1998)
2. Bhatt, R.B., Gopal, M.: On the compact computational domain of fuzzy-rough sets. Pattern Recognition Lett. 26, 1632–1640 (2005)
3. Düntsch, I., Gediga, G.: Uncertainty measures of rough set prediction. Artifi. Intell. 106, 109–137 (1998)
4. Hu, Q., Pedrycz, W., Yu, D., Lang, J.: Selecting discrete and continuous features based on neighborhood dicision error minimization. IEEE Trans. Syst. Man Cybern. Part B 40(1), 137–150 (2010)
5. Hu, Q., Xie, Z.X., Yu, D.R.: Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. Pattern Recognit. 40, 3509–3521 (2007)
6. Hu, Q., Yu, D.R., Xie, Z.X.: Information-preserving hybrid data reduction based on fuzzy-rough techniques. Pattern Recognit. Lett. 27(5), 414–423 (2006)
7. Hu, X.H., Cercone, N.: Learning in relational databases: a rough set approach. Int. J. Comput. Intell. 11(2), 323–338 (1995)
8. Jensen, R., Shen, Q.: Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches. IEEE Trans. Knowl. Data Eng. 16(12), 1457–1471 (2004)
9. Jensen, R., Shen, Q.: Fuzzy-rough sets assisted attribute reduction. IEEE Trans. Fuzzy Syst. 15(1), 73–89 (2007)
10. Jensen, R., Shen, Q.: New approaches to fuzzy-rough feature selection. IEEE Trans. Fuzzy Syst. 17(4), 824–838 (2009)
11. Liang, J.Y., Dang, C.Y., Chin, K.S., Yam Richard, C.M.: A new method for measuring uncertainty and fuzziness in rough set theory. Int. J. Gen. Syst. 31(4), 331–342 (2002)
12. Liang, J.Y., Qian, Y.H.: Information granules and entropy theory in information systems. Science in China Series F: Information Sciences 51(10), 1427–1444 (2008)
13. Liu, X.D., Pedrycz, W., Chai, T.Y., Song, M.L.: The development of fuzzy rough sets with the use of structures and algebras of axiomatic fuzzy sets. IEEE Trans. Knowl. Data Eng. 21(3), 443–462 (2009)
14. Mitra, S., Banka, H., Pedrycz, W.: Rough-fuzzy collaborative clustering. IEEE Trans. Syst. Man Cybern. Part B 36(4), 795–805 (2006)
15. Pawlak, Z.: Rough sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publisher, London (1991)
16. Pedrycz, W., Bargiela, A.: Granular clustering: a granular signature of data. IEEE Trans. Syst. Man Cybern. Part B 32(2), 212–224 (2002); Pedrycz, W., Kwak, K.: Boosting of granular models. Fuzzy Sets and Systems 157, 2934–2953 (2006)
17. Pedrycz, W., Sosnowski, Z.A.: Designing decision trees with the use of fuzzy granulation. IEEE Trans. Syst. Man Cybern. Part A 3(2), 151–159 (2000)

18. Qian, Y.H., Liang, J.Y., Pedrycz, W., Dang, C.Y.: Positive approximation: an accelerator for attribute reduction in rough set theory. Artifi. Intell. 174, 597–618 (2010)
19. Sen, D., Pal, S.K.: Generalized rough sets, entropy and image ambiguity measures. IEEE Trans. Syst. Man Cybern. Part B 39(1), 117–128 (2009)
20. Shannon, C.E.: The mathematical theory of communication. The Bell System Technical Journal 27(3-4), 373–423, 623-656 (1948)
21. Tang, E.C.C., Chen, D.G., Yeung, D.S., Wang, X.Z., Lee, J.W.T.: Attributes reduction using fuzzy rough sets. IEEE Trans. Fuzzy Syst. 16(5), 1130–1141 (2008)
22. Wang, X.Z., Tang, E.C.C., Zhao, S.Y., Chen, D.G., Yeung, D.S.: Learning fuzzy rules from fuzzy samples based on rough set technique. Inf. Sci. 177, 4493–4514 (2007)