An Improved Heuristic Minimal Attribute Reduction Algorithm Based on Condition Information Entropy

Baoyi Wang^{1, a}, Xuefei Li^{2, b}, Shaomin Zhang^{3, c}

¹School of Control and Computer Engineering, North China Electric Power University, Baoding, 071003, China

²School of Control and Computer Engineering, North China Electric Power University, Baoding, 071003, China

³School of Control and Computer Engineering, North China Electric Power University, Baoding, 071003, China

^aemail: wangbaoyiqj @126.com, ^bemail: lxf_182@163.com, ^cemail: zhangshaomin@126.com

Keywords: Rough set; Minimum attribute reduction; Entropy

Abstract. Attributes reduction was one of the key problems in rough set theory. However, to find the minimum attribute reduction is a NP-hard problem. This paper proposed an improved algorithm of heuristic minimal attribute reduction. First it calculates the attribute importance and gets core, then taking the core as a starting point, it selects attributes according to attribute importance which is defined by entropy, and gets the minimum attributes reduction. It's proved to be effective by theoretical analysis and example analysis.

Introduction

Rough Set theory is founded by a Polish mathematician called Pawlak[1]. Rough Set is a mathematics theory that can deal with incertitude and incompletion data, which has been successfully applied in data mining, pattern recognition, process control, information systems analysis, decision support systems, classification, and fault detection and other fields. Attributes reduction was one of the key problems in rough set theory. Alleged attributes reduction is deleting those redundant attribute which are not important or not related under the condition of the classification ability of information system.

The result of attributes reduction is always not unique, and to find the minimal attribute reduction is a NP-hard problem. Skoworon and Rauszer simplified discernibility matrix with absorption law and catch all the reductions, then find the the minimal attribute reduction[2], however their algorithm only can deal with decision table of 15 attributes. Srarzyk et al simplified discernibility matrix with extension rule and catch all the reductions[3], however his algorithm only can deal with decision table of 40 attributes. For the decision table of a large number of attributes, the algorithm complexity of time and space is beyond computer's capabilities. Chen YuMing proposed an algorithm of minimal attribute reduction based on power set tree in decision table[4], and the time complexity of this algorithm is $O(|U||C|2^{n-1})$, and the space complexity of this algorithm is $O(|U||C|2^{n-1})$.

In practice, one can only find the approximate minimal attributes reduction. Approximate algorithm is divided into two categories .One kind is minimal reduction algorithm based on stochastic optimization, such as the minimal reduction algorithm based on holographic particle swarm algorithm[5], the efficient combinatorial artificial bee colony algorithm for solving minimum attribute reduction problem[6], the minimum attribute reduction algorithm based on PSO[7], the knowledge reduction algorithm based on IPSO[8], the minimum rough set attribute reduction algorithm based on virus-coordinative discrete particle swarm optimization[9].The other kind is traditional heuristic minimal reduction algorithm. Lv Linxia proposed an algorithm of information system attribute reduction based on entropy[10], the time complexity of which is $O(|A|^3|U|^2)$. With this algorithm, computer can find the minimal attributes reduction or

approximate minimal attributes reduction. Li Longhui proposed an algorithm for the least attribute reduction of binary discernibility matrix [11], whose time complexity is $O(|C| \times |U/C|^2 \times |BM|)$ (BM is a binary discernibility matrix).

This paper introduces a new algorithm of attributes reduction. The algorithm used in the algorithm of equivalence class based on bucket sorting, with attribute importance information entropy as heuristic information. The algorithm is further improves the speed of finding approximate attributes reduction. The time complexity of the algorithm is analyzed in this paper, and the effectiveness of the proposed algorithm is verified by examples.

Relative Conceptions of Rough Set Theory

Definition 1[12] Information systems can be represented by a quadruple S = (U, A, V, f). In this quadruple, nonempty set $U = (x_1, x_2, \dots, x_n)$ is referred to as universe of discourse, A is a nonempty set of attributes, $V = \bigcup_{a \in A} V_a$, and V_a represents value region of attribute $a, f: U \times A \rightarrow V$ is a information function, that is $\forall x \in U, a \in A, f(x, a) \in V_a$.

Definition 2[12] Decision table is a special information system. In the quadruple S = (U, A, V, f), set A is made up of 2 parts, as $A = C \cup D$ and $C \cap D = \emptyset$. C is condition attributes set and D is decision attributes set. In general D only contains an attribute.

Definition 3[12] Regarding to information system S = (U, A, V, f), let $R \subseteq A$, then $ind(R) = \{(x_i, x_j) | f(x_i, b) = f(x_j, b), P_b \in R\}$ is referred to as indiscernibility relation of S. Obviously, indistinguishable relationship is an equivalence class, and the equivalence class which contains x is denoted as $[x]_R$. By U/R we denote the family of all equivalence class of R.

Definition 4[12] Regarding to information system S = (U, A, V, f), let X is a nonempty subset of U. Let $R \subseteq A$, then R lower approximation and R upper approximation of the set X is defined as follows:

$$\underline{R}X = \bigcup \{ Y \subseteq U / R \,|\, Y \subseteq X \} \tag{1}$$

$$\overline{R}X = \bigcup \{Y \subseteq U / R \mid Y \cap X \neq \emptyset\}$$
⁽²⁾

Definition 5[12] Regarding to decision table $S = (U, C \cup D, V, f), P \subseteq C$, P positive region of D is denoted as $POS_P(D)$, which is defined as $POS_P(D) = \bigcup_{X \in U/D} \underline{RX}$. Definition 6[12] Regarding to decision table $S = (U, C \cup D, V, f)$, $a \in C$, if

Definition 6[12] Regarding to decision table $S = (U, C \cup D, V, f)$, $a \in C$, if $POS_C(D) = POS_{C-\{a\}}(D)$, then a is unnecessary to C with respect to D. Otherwise, a is necessary to C with respect to D. The set of all attributes which are necessary is referred to as core of C with respect to D, and is denoted as Core(C).

Definition 7[13] Regarding to decision table $S = (U, C \cup D, V, f)$, let $U/C = \{X_1, X_2, \dots, X_m\}$, $U/D = \{Y_1, Y_2, \dots, Y_n\}$, then the conditional information entropy of the decision table is defined as follow:

$$H(D | C) = -\sum_{j=1}^{m} p(X_j) \sum_{i=1}^{n} p(Y_i | X_j) \log p(Y_i | X_j)$$
(3)

In the formula, $p(X_i) = \frac{|X_i|}{|U|}, i = 1, 2, \dots, n, p(Y_i | X_j) = \frac{|X_j \cap Y_i|}{|X_j|}, j = 1, 2, \dots, m, i = 1, 2, \dots, n.$

Theorem 1[13] Regarding to decision table $S = (U, C \bigcup D, V, f)$, $\forall a \in C$, the necessary and sufficient condition for that *a* is unnecessary to *C* with respect to *D* is $H(D | C) = H(D | C - \{a\})$.

Based on the concept of condition information entropy, the significance of attribute can be

analyzed from the perspective of condition information entropy.

Definition 8[13] Regarding to decision table $S = (U, C \cup D, V, f), \forall a \in C$, the significance to *C* of attribute *a* is denoted as $sgf_{C}(a)$, which is defined as $sgf_{C}(a) = H(D | C) - H(D | C - \{a\})$. Specially, when $C = \{a\}, sgf_{C}(a) = H(D | C) - H(D | \emptyset) = H(D | \{a\}), H(D | \emptyset) = 0$.

The significance to C of attribute a is determined by the change of condition information entropy without a. The more condition information entropy changes, the more significant a is.

Property 1 Attribute $a \in C$ is necessary to *C* with respect to *D* if and only if $sgf_C(a) > 0$. Property 2 $Core(C) = \{a \in C \mid sgf_C(a) > 0\}$

Theorem 2 Regarding to decision table $S = (U, C \cup D, V, f)$, $P \subseteq C$, if H(D | P) = H(D | C)and $\forall a \in P, sgf_{P-\{a\}}(a) > 0$, then P is a reduction of C.

An Algorithm of Heuristic Minimal Attribute Reduction

The significance of attribute reflect how significant attribute a is to condition attributes set C.Using the idea of greedy algorithm, this paper proposes an algorithm of heuristic minimal attribute reduction. First it calculates the attribute importance and gets core, then taking the core as a starting point, it selects attributes according to attribute importance which is defined by entropy, and gets the minimum attributes reduction.

In the algorithm, the condition information entropy is calculated repeatedly, so here is introduced an algorithm of equivalence class partition based on bucker sort.

```
Algorithm 1 Input U = \{u_1, u_2, \dots, u_n\}, P \subset C, D, output H(D | P)
GetH(U.P.D)
{
   let H=0;
   calculate equivalence class partition of D: U' = U / P = \text{get\_ind}(U,D);
   while (U' \neq \emptyset)
   {
     select a element u' from set U', let U' = U' - \{u'\};
     calculate p = \frac{|u'|}{|U|};
     calculate equivalence class partition of P: U'' = u' / P = \text{get\_ind}(u', P);
     h = 0;
     while (U'' \neq \emptyset)
     {
       select a element u'' from set U'', let U'' = U'' - \{u'\};
       calculate p' = \frac{|u''|}{|u'|};
        h = h + p' \times \log p';
     H = H - p \times h;
   return H;
}
```

In the algorithm , get_ind(U,P) is an algorithm of equivalence class partition based on bucker sort proposed in paper[14]. It's time complexity is O(|P||U|). Behind calculating U' = U/P, there are 2 circulations in the algorithm 1. The outer circulation contains get_ind(u', P) and inner circulation. Obviously, the time complexity of inner circulation is $O(u'_i/P)$, so the time complexity of one time outer circulation is $O(|P||u'_i|) + O(|u'_i/P|) = O(|P||u'_i|)$. Let $U' = \{u'_1, u'_2, \dots, u'_n\}, U'$

meets 2 properties as follow:

 $u'_{i} \cap u'_{j} = \emptyset, \quad i, j = 1, 2, \dots, i \neq j$ $u'_{1} \cup u'_{2} \cup \dots \cup u'_{n} = U'$ (4) (5)

So, the time complexity of outer circulation is $\sum_{i=1}^{n} O(|P||u'_i|) = O(|P| \times (|u'_1| + |u'_2| + \dots + |u'_n|)) = O(|P||U'|),$

```
which is O(|P||U|) for the worst case.
```

```
Algorithm 2 Input decision table S = (U, C \bigcup D, V, f), output minimal reduction.
GetMinReduction(S)
{
   calculate H(D | C);
   build set Core(C) = \emptyset;
   for each a \in C
   {
      calculate sgf_{C}(a) = H(D | C) - H(D | C - \{a\});
      if (sgf_c(a) > 0)
           add attribute a to Core(C);
   }
   let C' = Core(C);
   while (H(D | C') \neq H(D | C))
   {
      for each a \in C - C' calculate sgf_{C'+\{a\}}(a) = H(D | C' + \{a\}) - H(D | C');
      select a' \in C - C' which meets sgf_{C'+\{a'\}}(a) = \max_{a \in C'} sgf_{C'+\{a\}}(a), and add it to C';
   }
   return C';
}
```

The Algorithm Complexity Analysis

The algorithm is mainly divided into three steps, as follows:

- 1) calculate H(D|C);
- 2) calculate core of attributes set C;
- 3) according to the significance of attribute, select attribute and add to reduction set.

The time complexity of step 1 is O(|C||U|), which is stated in the previous section.

In step 2, there is a circulation. This circulation executes |C| times, and each execution needs to calculate $H(D | C - \{a\})$. So the time complexity of this circulation is O((|C|-1)|U|) = O(|C||U|), and the time complexity of step 2 is $O(|C|^2|U|)$.

In step 3, there is 2 circulations. The time of executions of the outer circulation is not more than |C|. The time of executions of the inner circulation is |C-C'|, which is |C| for the worst case. Each execution needs to calculate $H(D|C'+\{a\})$, so it's time complexity is O((|C'|+1)|U|) = O(|C'||U|), which is O(|C||U|) for the worst case. So the time complexity of step 3 is $O(|C| \times |C| \times |C| \|U|) = O(|C|^3 |U|)$.

From what has been discussed above, the time complexity of the algorithm of heuristic minimal attribute reduction is $O(|C||U|) + O(|C|^2|U|) + O(|C|^3|U|) = O(|C|^3|U|)$. The time complexity of the algorithm proposed in paper[10] is $O(|A|^3|U|^2)$. The time complexity of the algorithm proposed in paper[11] is $O(|C| \times |U/C|^2 \times |BM|)$. BM presents binary discernibility matrix, and $|BM| = |U|^2$, so It's actual time complexity is $O(|C| |U/C||U|^2)$. Compared with those two

algorithms, the algorithm in this paper is faster.

Example Analysis

For example, in decision table $S = (U, C \cup D, V, f)$, $U = \{u_1, u_2, \dots, u_8\}$, $C = \{c_1, c_2, \dots, c_6\}$, $D = \{d\}, V_x = \{1, 2, 3, 4\}$ $x = c_1, c_2, \dots, c_6, d$, f is stated as table 1, please find the minimal reduction. Table 1 Decision table

Table 1 Decision table						
	c ₁	c_2	c ₃	c_4	c ₅	d
u ₁	3	1	2	3	4	4
u ₂	3	2	3	3	1	3
u ₃	3	2	2	3	4	4
u_4	2	3	1	3	2	4
u ₅	2	3	1	3	2	2
u ₆	3	3	2	3	4	4
u ₇	3	2	3	3	1	3
u ₈	2	1	1	3	3	2

Firstly, calculate H(D|C). $U/C = \{\{u_1\}, \{u_2, u_7\}, \{u_3\}, \{u_4, u_5\}, \{u_6\}, \{u_8\}\}, \{u_8\}\}$

 $U/D = \{\{u_1, u_3, u_4, u_6\}, \{u_2, u_7\}, \{u_5, u_8\}\}, H(D | C) = 1.25$

Secondly, calculate core of C. $sgf_C(c_2) = 0.59$, $sgf_C(c_i) = 0$ i = 1,3,4,5, so $Core(C) = \{c_2\}$. Thirdly, select the most significant attribute and add to reduction set. Let C' = Core(C), $sgf_{C'+\{c_i\}}(c_i) = 0$ i = 1,3,4, $sgf_{C'+\{c_5\}}(c_5) = 1.25$, so add c_5 to C'. then H(D | C') = 1.25 = H(D | C), so $C' = \{c_2, c_5\}$ is a reduction of C to D, and the minimal reduction of C to D.

This result is consistent with the result calculated by discernibility matrix.

Conclusion

This paper proposed an improved algorithm of heuristic minimum attribute reduction. First it calculates the attribute importance and gets core, then taking the core as a starting point, it selects attributes according to attribute importance which is defined by entropy, and gets the minimal attributes reduction. This paper analyzes it's time complexity, $O(|C|^3|U|)$. It's proved to be effective by example analysis.

References

[1] Pawlak Z,Slowinski R. Rough set [J]. Int J of Computer and Information Science, 1982, 11(5):341-356

[2] Skowron A, Rauszer C. The discernibility matrices and functions in information system[C]. Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory. Dordrecht: Kluwer AcademicPublishers,1992:331-362.

[3] Starzyk J, NeIson D E, Sturtz K.Reduct generation in information system[J]. Bulletin of International Rough Set Society,1999;3(1/2):19~22.

[4] Chen Yu-Ming, Wu Ke-Shou, Sun Jin-Hua. Minimal attribute reduction based on power set tree in decision table[J]. Journal Of Nanjing University(Natural Sciences), 2012, 02:164-171.

[5] MA Sheng-Lan,YE Dong-Yi. Research on Computing Minimum Entropy Based Attribute Reduction via Stochastic Optimization Algorithms [J]. Pattern Recognition and Artificial Intelligence, 2012, 25(1): 96-104.

[6] YE Dong-Yi,CHEN Zhao-jiong. An Efficient Combinatorial Artificial Bee Colony Algorithm for Solving Minimum Attribute Reduction Problem [J]. Acta Electronica Ainica, 2015, 43(5): 1014-1020.

[7] YANG Xiao-yan, CHEN Guo-long, GUO Wen-zhong. Minimum attribute reduction algorithm based on particle swarm optimization [J]. Journal of Fuzhou University(Natural Science), 2010, 38 (2): 193-197.

[8] HUANG Fu-yuan . A Knowledge Reduction Algorithm Based on IPSO [J]. Modern Computer(professional),2013,31:6-9.

[9] TAO Xin-min, WANG Yan, XU Jing. Minimum rough set attribute reduction algorithm based on virus-coordinative discrete particle swarm optimization [J]. Control andDecesion, 2012, 27(2): 259-265.

[10]Lv Linxia, Zhao Xiying, Tang Zhanhong. An algorithm of information system attribute reduction based on entropy[J].Automation and instrumentation,2013,05:197-199.

[11] LI Long-shu, WANG Hui-ping ,XU Yi. Algorithm for the Least Attribute Reduction of Binary Discernibility Matrix [J].Computer Technology and Development, 2010, 20(6):93-100.

[12] WANG Guo-Yin.Rough Set Theory and Knowledge Acquisition[M]. XI'AN:XI'AN JIAOTONG UNIVERSITY PRESS,2001.5.

[13] WANG Guo-Yin, YU Hong, YANG Da-Chun. Decision Table Reduction based on Conditional Information Entropy [J]. CHINESE JOURNA L OF COMPUTERS, 2002, 25(07):759-766.

[14] JIANG Yu, LIU Yin-tian, LI Chao. Fast algorithm for computing attribute reduction based on Bucket Sort [J]. Control and Decision,2011,26(02):207-212.