Knowledge-Based Systems 102 (2016) 116–126

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Discrete particle swarm optimization approach for cost sensitive attribute reduction

Jianhua Dai[a,b,*], Huifeng Han[b], Qinghua Hu[a], Maofu Liu[c]

[a] School of Computer Science and Technology, Tianjin University, Tianjin, 300350, China
[b] College of Computer Science and Technogloy, Zhejiang University, Hangzhou, 310027, China
[c] College of Computer Science and Technology, Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan, 430065, China

ARTICLE INFO

Article history:
Received 14 October 2015
Revised 31 March 2016
Accepted 2 April 2016
Available online 5 April 2016

Keywords:
Attribute reduction
Rough set theory
Cost sensitive attribute reduction
Particle swarm optimization

ABSTRACT

Attribute reduction is a key issue in rough set theory which is widely used to handle uncertain knowledge. However, most existing attribute reduction approaches focus on cost insensitive data. There are relatively few studies on cost sensitive data. Especially, how to evaluate a cost sensitive reduction algorithm is still an issue needing to be studied further. In this paper, we propose four relative evaluation metrics which can be used to compare and evaluate different algorithms for cost sensitive attribute reduction more conveniently. Moreover, we propose a particle swarm optimization method for cost sensitive attribute reduction problem inspired by its powerful search ability. The proposed approach is tested with three typical test cost distributions and compared with an influential algorithm reported recently on both exiting metrics and proposed metrics. Results indicate that the proposed relative evaluation metrics are effective and convenient. Comparing results also show that the proposed algorithm is effective.

1. Introduction

Rough set theory is a useful tool for analysis and processing of data [28,29], and has draw great interests in both theoretical and application aspects [6,10–14,16,20,22,28,36–39]. Using the concepts of lower and upper approximations, rough set theory can obtain hidden knowledge from information systems or information tables. Furthermore, it can also be used to achieve a subset of attributes of original information systems, called an attribute reduct. Actually, an attribute reduct is a subset of attributes that are jointly sufficient and individually necessary for preserving a particular property of a given information table. In recent years, numerous studies on attribute reduction based on rough set theory have been proposed [8,9,15,17,30,31,35,40]. Please refer to the review paper [33] for more information about attribute reduction in rough set theory.

Most existing attribute reduction approaches deal with cost insensitive data. In many applications, however, data are not free. Each test (i.e. attribute, feature) may have an associated cost [23,25,26,32,34]. Test cost is the money, time, or other resources we pay for collecting a data item of an object [25,34]. For example, in medical diagnosis, a blood test has a cost which may be quite different to the cost of an fMRI test.

Min et al. [24] addressed the feature selection with test cost constraint problem. Susmaga [32] introduced an exhaustive algorithm for generating all reducts of the minimal cost, called the minimal cost reducts or the cheapest reducts. Min et al. [23] studied the similar problem, called the minimal test cost reduct problem, and proposed a heuristic algorithm. They also defined three metrics to evaluate the performance of reduction algorithms. The metrics supply measures to evaluate algorithms for the minimal test cost reduct problem. However, there is an assumption that all the reducts are obtained when computing these metrics. In other words, one needs to obtain all the reducts, which is really a hard task, to compute the metrics. To our understanding, existing metrics for comparing algorithms are a bottleneck for constructing other new algorithms. Hence, we need to construct metrics which are not defined based on the global optimal reduct.

Particle swarm optimization (PSO) is a kind of swarm intelligence optimization method based on social-psychological principles and provides insights into social behavior, as well as contributing to engineering applications. Kennedy and Eberhart first proposed PSO algorithm [21]. PSO provides potential solutions through particles flying across the problem hyperspace. The unique information diffusion and interaction mechanisms of PSO enable it to solve many problems with good performance at low

* Corresponding author.
E-mail address: david.joshua@qq.com (J. Dai).

http://dx.doi.org/10.1016/j.knosys.2016.04.002
0950-7051/

computational cost. PSO has been successfully applied into many fields [1,4,5,18,19,27]. Inspired by the powerful search ability, Dai et al. introduced discrete PSO into attribute reduction problem [7]. However, cost factor was not considered in [7].

In this paper, we focus on the minimal test cost reduct problem including the evaluation metrics and the computing of the minimal test cost reduct. To compare and evaluate different algorithms for the minimal test cost reduct problem, we propose four new relative metrics. These metrics are convenient to compute. Moreover, we construct a particle swarm optimization method for cost sensitive attribute reduction. One reason why we choose PSO is because of the fact that PSO is mainly used to handle function optimization rather than combination optimization. It is very convenient to use PSO to handle function optimization problems, and is not suitable for combination optimization problems directly. The minimal cost attribute reduct problem is a typical combination optimization problem. We hope that other search algorithms, such as Genetic Algorithm, Genetic Programming, Artificial Immune Algorithm, Ant Colony Algorithm, Artificial Bee Colony Algorithm, can be easily used to the minimal cost attribute reduct problem even if some of them were first proposed to handle function optimization problems based on our study. The proposed algorithm based on PSO is evaluated by existing three metrics and the proposed four metrics and compared with the existing heuristic algorithm in [23]. Results indicate the effectiveness of the proposed metrics and algorithm.

The rest of this paper is organized as follows. In Section 2, the minimal test cost attribute reduct problem and related concepts are reviewed. In Section 3, existing evaluation metrics are reviewed. Four new evaluation metrics are proposed. Section 4 presents a method based on particle swarm optimization for the minimal test cost attribute reduct problem. In Section 5, experiments are conducted and the comparing results are shown. Section 6 concludes this paper.

2. Cost sensitive attribute reduction

In this section, we review some definitions related to cost sensitive attribute reduction problem.

Definition 2.1 [23]**.** A test-cost-independent decision system (TCI-DS) S is the 6-tuple:

$$S =< U, C, D, \{V_a | a \in C \cup D\}, \{I_a | a \in C \cup D\}, c > \tag{1}$$

where U is a finite set of objects called the universe, C is the set of conditional attributes, D is the set of decision attributes, $\{V_a\}$ is the set of values for each $a \in C \cup D$, and $\{I_a\}$: $U \rightarrow V_a$ is an information function for each $a \in C \cup D. c : C \rightarrow R^+ \cup \{0\}$ is the test cost function and

$$c^*(A) = \sum_{a \in A} c^*(\{a\}) = \sum_{a \in A} c(a) \tag{2}$$

where $c^* : 2^C \rightarrow R^+ \cup \{0\}$ is the attribute subset test cost function.

Definition 2.2. The positive region of D with respect to $B \subseteq C$ is defined as:

$$POS_B(D) = \bigcup_{X \in U/D} \underline{B}(X) \tag{3}$$

where $\underline{B}$ (X) denotes the B-lower approximation of X, and

$$\underline{B}(X) = \{x_i \in U | [x_i]_B \subseteq U\} \tag{4}$$

where $[x_i]$ is the equivalent class containing x under the indiscernibility relation generated by B.

Definition 2.3. Any $B \subseteq C$ is called a relative reduct of S if and only if:

(i) $POS_B(D) = POS_C(D)$
(ii) $\forall a \in B, POS_{B-\{a\}} \subset POS_C(D)$

Definition 2.4 [23]**.** Let $Red(S)$ denote the set of all relative reducts of a test-cost-sensitive decision system S. Any $R \in Red(S)$ where $c^*(R) = min\{c^*(R') | R' \in Red(S)\}$ is called a minimal test cost reduct.

There may exist a number of minimal test cost reducts. The minimal test cost reduct problem is to find any one of them.

3. Cost setting and evaluation metrics

Three different schemes to produce random test costs were employed in [23]. In this paper, we use the same schemes which comprise: uniform distribution, normal distribution, and Pareto distribution. For simplicity, test costs are integers ranging from M to N, and are evaluated independently.

(a) Uniform distribution

This is one of the most commonly used distribution. Let c denote the cost value of an attribute, then

$$P(c = n) = \frac{1}{N - M + 1} \quad \text{where} \quad n \text{ is an integer in } [M, N].$$

This can be generated from a uniformly distributed random number x on (0,1).

$$c_u(M, N, x) = M + \lfloor (N - M + 1)x \rfloor$$

(b) Normal distribution

Normal distribution is defined as:

$P(c) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(c-u)^2}{2\sigma^2}}$ where parameters u and σ^2 are the mean and the variance respectively.

Like in [23], to generate an integer in [M, N], we first obtain a random number r from a normal distribution with $u = 0$ and $\sigma = 1$. Then, let $y = \frac{M+N+1}{2} + \alpha r$, where α is a real number which equals to 8.

Clearly, y is also a normal distribution with $u = \frac{M+N+1}{2}$.

Finally, we set the cost as follows:

$$\mathrm{c_n(M, N, r)} = \begin{cases} M; & y < M \\ N; & y > N \\ \lfloor y \rfloor; & otherwise \end{cases} \tag{5}$$

(c) Pareto distribution

Pareto distribution is a power probability distribution that is used in description of social, scientific, geophysical, actuarial, and many other types of observable phenomena.

The bounded Pareto distribution which is more applicable is used, similar to [23]. For a uniformly random number x in (0,1), we have

$$P(M, N, x) = \left(-\left(\frac{x(N+1)^\alpha - xM^\alpha - (N+1)^\alpha}{M^\alpha (N+1)^\alpha} \right) \right)^{\frac{1}{\alpha}}$$

It is a bounded Pareto-distribution on (M,N + 1) with α determines the shape of the distribution. In our experiment, α is set to 2.

Finally, we set the cost $c_p(M, N, x) = \lfloor P(M, N, x) \rfloor$.

To evaluate the performance of an algorithm for the cost sensitive attribute reduct problem, Min et al. [23] defined three metrics including *finding optimal factor, maximal exceeding factor* and *average exceeding factor*.

Definition 3.1 [23]**.** Let the number of experiments be K, and the number of successful searches of an optimal reduct be k. The *Finding Optimal Factor* (FOF) is defined as:

$$op = \frac{k}{K} \tag{6}$$