Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 946070, 12 pages
http://dx.doi.org/10.1155/2013/946070

Research Article

Test-Cost-Sensitive Attribute Reduction of Data with Normal Distribution Measurement Errors

Hong Zhao, Fan Min, and William Zhu

Laboratory of Granular Computing, Zhangzhou Normal University, Zhangzhou 363000, China

Correspondence should be addressed to Fan Min; minfanphd@163.com

Received 31 December 2012; Accepted 1 March 2013

Academic Editor: Hung Nguyen-Xuan

The measurement error with normal distribution is universal in applications. Generally, smaller measurement error requires better instrument and higher test cost. In decision making, we will select an attribute subset with appropriate measurement error to minimize the total test cost. Recently, error-range-based covering rough set with uniform distribution error was proposed to investigate this issue. However, the measurement errors satisfy normal distribution instead of uniform distribution which is rather simple for most applications. In this paper, we introduce normal distribution measurement errors to covering-based rough set model and deal with test-cost-sensitive attribute reduction problem in this new model. The major contributions of this paper are fourfold. First, we build a new data model based on normal distribution measurement errors. Second, the covering-based rough set model with measurement errors is constructed through the "3-sigma" rule of normal distribution. With this model, coverings are constructed from data rather than assigned by users. Third, the test-cost-sensitive attribute reduction problem is redefined on this covering-based rough set. Fourth, a heuristic algorithm is proposed to deal with this problem. The experimental results show that the algorithm is more effective and efficient than the existing one. This study suggests new research trends concerning cost-sensitive learning.

1. Introduction

The measurement error is the difference between a measurement value and its true value. It can come from the measuring instrument, from the item being measured, from the environment, from the operator, and from other sources [1]. As a plausible distribution for measurement errors, the normal distribution was put forward by Gauss in 1809. In fact, normal distribution is found to be applicable over almost the whole of science and engineering measurement. In data mining applications, the data model based on measurement errors is an important form of uncertain data (see, e.g., [2–4]).

Test costs refer to time, money, or other resources spent in obtaining data items related to some object [5–10]. There are a number of measurement methods with different test costs to obtain a data item. Generally, higher test cost is required to obtain data with smaller measurement error. In data mining applications, we will select an attribute subset with appropriate measurement error to minimize the total test cost and at the same time preserve necessary information of the original decision system.

An attribute reduct is a subset of attributes that are jointly sufficient and individually necessary for preserving a particular property of the given information table [11]. It is a key problem of rough set theory and has attracted much attention in recent years (see, e.g., [12–16]). As a generalization of attribute reduction, test-cost-sensitive attribute reduction [6] focuses on selecting a set of tests to satisfy a minimal test cost criterion.

Recently, error-range-based covering rough set [4] was introduced to address error ranges. This theory is based on both covering-based rough set [17–23] and neighborhood rough set [24–28]. Moreover, in the new theory, the test-cost-sensitive attribute reduction problem deals with numeric data instead of nominal ones. Therefore, the problem is more challenging than the one defined in [6]. However, error-range-based covering rough set considers only uniform distribution errors, which are rather unrealistic.

In this paper, we introduce normal distribution to build a new model of covering-based rough set to address normal distribution measurement errors (NDME) according to the "3-sigma" rule. The major contributions of this paper are fourfold. First, we introduce normal distribution to build a new data model based on measurement errors. The error range is computed according to the values of attributes instead of the fixed error range for different datasets. Second, we build the computational model, namely, the covering-based rough set with normal distribution measurement errors. Third, the minimal test cost attribute reduction problem is redefined in the new model. Fourth, we propose a heuristic algorithm to address the reduction problem. Specifically, a δ-weighted heuristic reduction algorithm is designed, where attribute significance is adjusted by δ-weighted test cost.

Ten open datasets from the University of California-Irvine (UCI) library are employed to study the performance and effectiveness of our algorithm. We adopt three measures to evaluate the performance of the reduction algorithms from a statistical viewpoint. Experiments undertaken with open source software cost-sensitive rough sets (Coser) [29] validate the performance of this algorithm. Experimental results show that our algorithm can generate a minimal test cost reduct in most cases. At the same time, the proposed algorithm can achieve better performance and efficiency than the existing one [4].

The rest of the paper is organized as follows: Section 2 presents the data models with measurement errors and test costs, respectively. Section 3 describes the computational model, namely, covering-based rough set model with normal distribution measurement errors. The minimal test cost reduct problem under the new model is also defined in this section. Next, Section 4 presents a δ-weighted heuristic reduction algorithm and a competition approach. Experiment results and comparison with the existing work are discussed in Section 5. Finally, conclusions are drawn in Section 6.

2. Data Models

This section presents data models. First, we propose a decision system with normal distribution measurement errors, which is also called NEDS for brevity. Then, we introduce test costs to NEDS and define test-cost-sensitive decision systems with NDME.

2.1. Normal Distribution. Normal distribution is an important type in science and engineering measurement. It can be described by the probability density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \tag{1}$$

where the parameter μ is the mean which gives the location of the distribution and the parameter σ^2 is the variance which gives the scale of the distribution.

The cumulative distribution function (CDF) $F(x)$ of a random variable is the probability of a value less than or equal to some value x:

$$F(x) = \int_{-\infty}^{x} f(x)\, dx, \tag{2}$$

where $x \in \mathbb{R}$. For a random variable X,

$$F(x) = Pr(X \le x), \tag{3}$$

where the right-hand side represents the probability that the random variable X takes on a value less than or equal to x. The standard normal distribution appears with $\mu = 0$ and $\sigma^2 = 1$. The equation becomes

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{(-z^2)/2}. \tag{4}$$

For a normal distribution, about 68.27% of values drawn from a normal distribution are within one standard deviation away from the mean; about 95.45% of the values lie within two standard deviations; nearly all values (99.73%) lie within 3 standard deviations of the mean, that is, "3-sigma" rule [30]. We use the following example to explain the relationship between standard deviation and confidence interval.

Example 1. Let standard deviation be 0.01, and let the mean be 0; then we know that about 99% of the measurement errors are from −0.03 to +0.03.

2.2. Decision Systems with Measurement Errors. We introduce normal distribution measurement errors into our model [31] to make the model more realistic.

Definition 2. A decision system with normal distribution measurement errors (NEDS) S is the 6-tuple:

$$S = \left(U, C, D, V = \{V_a \mid a \in C \cup D\},\ I = \{I_a \mid a \in C \cup D\}, n\right), \tag{5}$$

where U is the nonempty set called a universe and C and D are the nonempty sets of variables called conditional attributes and decision attributes, respectively. V_a is the set of values for each $a \in C \cup D$, and $I_a : U \to V_a$ is an information function for each $a \in C \cup D$. $n : C \to \mathbb{R}^+ \cup \{0\}$ is the maximum value of measurement error. $+n(a)$ and $-n(a)$ are the upper confidence limit (UCL) and the lower confidence limit (LCL) of $a \in C$, respectively.

Definition 3. Letting $S = (U, C, D, V, I, n)$ be a NEDS, the error range of attribute a is defined as

$$n(a) = \Lambda e(a), \tag{6}$$

where

$$e(a) = \Delta \frac{\sum_{i=1}^{m} a(x_i)}{m}, \tag{7}$$

where $\Delta \in [0, 1]$ is a user-specified parameter, $a(x_i)$ is the ith instance value of $a \in C$, $i \in [1, m]$, and m is the number of instances. The precision of $e(a)$ can be adjusted through Δ setting.

From Definition 3 we can see that the decision system with normal distribution measurement errors is a generalization of the decision system and the decision system with error range (DS-ER) (see, e.g., [4]). If all attributes are error free, that is $\Lambda = 0$, a NEDS degrades to a DS. If the error range is a fixed value, that is $\Lambda = 1$, a NEDS degrades to a DS-ER.

We introduce how to deal with the abnormal value of measurement error. In applications, if the repeated measurement data satisfy

$$\left|x_i - \overline{x}\right| > 3\sigma, \quad (i = 1, 2, \ldots, N), \tag{8}$$

the x_i would be considered as an abnormal value and be rejected, where x_i is the ith measurement value and $\overline{x}$ is the mean of all measurement values. This is the Pauta criterion of measurement error theory.

Now, we investigate the relationship between the limit of confidence interval and the standard deviation in the following proposition.

Proposition 4. *Let $-n(a)$ and $+n(a)$ be LCL and UCL, respectively, and let Pr be the confidence level. One has the upper limit of confidence interval*

$$n(a) = 3\sigma, \tag{9}$$

where Pr = 99.73%.

The value of exceed the three deviations is an abnormal error, which needs to be identified and removed from consideration. The standard normal distribution is a special case of the normal distribution. The limit of confidence interval is investigated in the following proposition.

Proposition 5. *Let $-n(a)$ and $+n(a)$ be LCL and UCL of standard normal distribution measurement errors, respectively. One has*

$$n(a) = 3. \tag{10}$$

Proof. The standard normal distribution is given by taking $\mu = 0$ mean and $\sigma^2 = 1$ in a general normal distribution. $n(a) = 3\sigma$, $n(a) > 0$. Therefore (10) holds. □

In Definition 3, a key parameter Λ is an adjusting factor. Now we introduce it by the following proposition.

Proposition 6. *Let $-n(a)$ and $+n(a)$ be LCL and UCL of $a \in C$, respectively. Confidence intervals are stated at the Pr confidence level, and $n(a) = 3\sigma$. According to (3), one has*

$$Pr(|x| \le n(a)) = 2F(n(a)) - 1. \tag{11}$$

According to (3) and Proposition 6, if $2/3 \le \Lambda < 1$, we have $2\sigma \le n(a) < 3\sigma$, $95.45\% \le Pr < 99.73\%$; if $1/3 < \Lambda < 2/3$, we have $\sigma < n(a) < 2\sigma$, $68.27\% < Pr < 95.45\%$, and if $0 < \Lambda \le 1/3$, we have $0 < n(a) \le \sigma$, $0\% < Pr \le 68.27\%$.

Large error ranges are pronounced with shorter reaction time than those with smaller error ranges. Small error ranges are pronounced with higher classification accuracies than those with larger ones. Generally, smaller measurement error requires better instrument and higher test cost. In many applications, it is impossible or unnecessary to distinguish objects or elements with small error range in a universe. One can adjust the size of the error range through the Λ setting to meet different requirements.

Table 1: An example of numerical value attribute decision table.

Plant	SL	SW	PL	PW	Class
x_1	0.23529	0.77273	0.14286	0.04762	Setosa
x_2	0.29412	0.72727	0.11905	0.04762	Setosa
x_3	0.35294	0.09091	0.38095	0.42857	Versicolor
x_4	0.64706	0.31818	0.52381	0.52381	Versicolor
x_5	0.41176	0.31818	0.50000	0.42857	Versicolor
⋮	⋮	⋮	⋮	⋮	⋮
x_{149}	0.58824	0.54545	0.85714	1.00000	Virginica
x_{150}	0.44118	0.27273	0.64286	0.71429	Virginica

2.3. Test-Cost-Independent Decision System with Normal Distribution Measurement Errors. We introduce test costs to the data model. Now, we discuss the new model as follows.

Definition 7. A test-cost-independent decision system with normal distribution measurement errors (TCI-NEDS) S is the 7-tuple:

$$S = (U, C, D, V, I, n, c), \tag{12}$$

where U, C, D, V, I, and n have the same meanings as in a NEDS, and $c : C \to \mathbb{R}^+ \cup \{0\}$ is the test cost function. Test costs are independent of one another, that is, $c(B) = \sum_{a \in B} c(a)$ for any $B \subseteq C$.

Note that in this model, test costs are not applicable to decision attributes.

In order to process and compare, the values of conditional attributes are normalized from their value into a range from 0 to 1 through the linear function

$$y = \frac{(x - \min)}{(\max - \min)}, \tag{13}$$

where max and min are the maximal and minimal values of the attribute and x and y are the initial value and the normalized value, respectively.

Table 1 presents a decision system of Iris, whose conditional attributes are normalized values. Where $C = \{\text{SL, SW, PL, PW}\}$, $D = \{\text{Class}\}$, and $U = \{x_1, x_2, \ldots, x_{150}\}$.

3. Covering-Based Rough Set with Normal Distribution Measurement Errors

Rough set theory is a powerful tool for dealing with uncertain knowledge in information systems [32]. It has been successfully applied to feature selection [33, 34], rule extraction [35–37], uncertainty reasoning [38, 39], decision evaluation [40, 41], granular computing [42–45], and so forth. Recently,

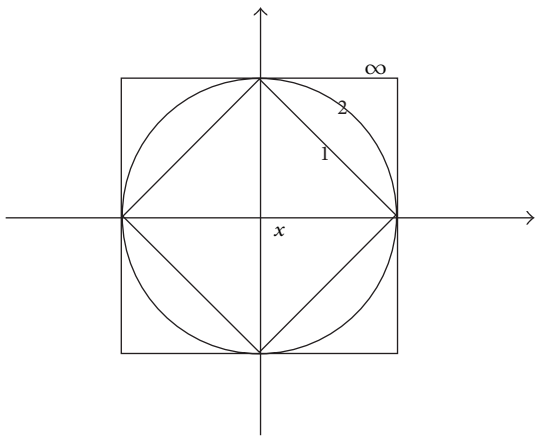

Figure 1: Conventional neighborhoods.

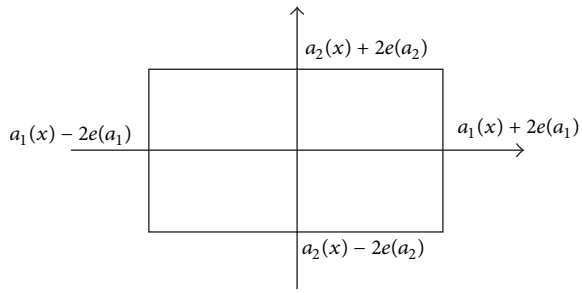

Figure 2: Two-dimensional neighborhood with error ranges.

covering-based rough set has attracted much research interest with significant achievements in both theory and application.

The concept of neighborhood (see, e.g., [46–48]) has been applied to define different types of covering-based rough set [16–18]. From the different viewpoints, a neighborhood is called an information granule or a covering element. Figure 1 illustrates the neighborhoods of x in a two-dimensional real space [25]. For this type of neighborhood rough set model, the distance parameter δ is a user-specified parameter. Objects with a distance less than δ are viewed as neighbors. Recently, a new neighborhood is defined in [4]. The size of the neighborhood depends on error ranges of tests, and more objects fall into the neighborhood of x_i for wider error ranges. Figure 2 illustrates this two-dimensional neighborhood.

In this section, we introduce normal distribution measurement errors to covering-based rough set. The new model is called covering-based rough set with normal distribution measurement errors. As mentioned early, if we do not consider errors, this mode degenerates to the classical decision system. Therefore, this model is a natural extension of classical one. Test-cost-sensitive attribute reduction problem on the covering-based rough set model with NDME is also proposed in this section.

3.1. Covering-Based Rough Set with Normal Distribution Measurement Errors. According to "3-sigma" rule, we present a new model considering both error distribution and confidence interval. According to Definition 2, we have defined a neighborhood in [31] as follows.

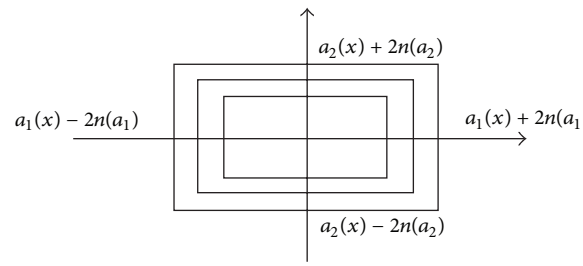

Figure 3: Two-dimensional neighborhood based on different standard deviations.

Definition 8. Let $S = (U, C, D, V, I, n)$ be a decision system with normal distribution measurement errors. Given $x_i \in U$ and $B \subseteq C$, the neighborhood of x_i with respect to normal distribution measurement errors on attribute set B is defined as

$$n_B(x_i) = \{x \in U \mid \forall a \in B, |a(x) - a(x_i)| \leq 2n(a)\}, \quad (14)$$

where $n(a) = \Lambda e(a)$ is the error boundary. It represents the error value of a in $[-n(a), +n(a)]$.

Measurement errors with no more than a difference of $2n(a)$ should be viewed as the family of neighborhood granules. We explain why $n(a)$ instead of $e(a)$ was employed in (14) as the maximal distance. Although the value of error is within a certain range, there are significant differences among confidence intervals. As mentioned earlier, "3-sigma" rule states that for a normal distribution, different proportion values lie within different standard deviations of the mean. In particularly, the proportion is very close to 0 if data is more than three standard deviations from the mean.

Therefore, measurement errors with no more than a difference of $n(a) = \Lambda e(a)$ should be viewed as the family of neighborhood granules. Naturally, the size of the neighborhood depends on error ranges of tests and adjusting factor. Figure 3 shows the different sizes of neighborhood based on $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$.

In the new model, the lower and upper approximations are defined as follows.

Definition 9 (see [31]). Let $S = (U, C, D, V, I, n)$ be a decision system with normal distribution measurement errors, and let N_B be a neighborhood relation induced by $B \subseteq C$. For any $X \subseteq U$, the lower and upper approximations of X in a neighborhood approximation space $\langle U, N_B \rangle$ are defined as

$$\underline{N_B}(X) = \{x_i \mid x_i \in U \wedge n_B(x_i) \subseteq X\},$$
$$\overline{N_B}(X) = \{x_i \mid x_i \in U \wedge n_B(x_i) \cap X \neq \emptyset\}. \quad (15)$$

$\forall X \subseteq U$, $\overline{N_B}(X) \supseteq X \supseteq \underline{N_B}(X)$. The boundary region of X in $\langle U, N_B \rangle$ is defined as

$$BN_B(X) = \overline{N_B}(X) - \underline{N_B}(X). \quad (16)$$

The positive region of D with respect to $B \subseteq C$ is defined as $\mathrm{POS}_B(D) = \bigcup_{X \in U/D} \underline{N_B}(X)$ [49, 50].

3.2. Test-Cost-Sensitive Attribute Reduct Problem. Attribute reduction is a successful technique to remove redundant data and facilitate the mining task. A number of definitions of relative reducts exist [25, 38, 51, 52] for different rough set models. In this section, we define test-cost-sensitive attribute reduction on the covering-based rough set model with NDME.

A minimal test cost reduct problem proposed in [6] can be redefined as follows. The problem of finding such a reduct is called the *minimal test cost reduct problem*.

Problem 1. The minimal test cost reduct problem.

Input: $S = (U, C, D, V, I, n, c)$;

Output: $B \subseteq C$;

Constraint: $\mathrm{POS}_B(D) = \mathrm{POS}_C(D)$;

Optimization objective: $\min(c(B))$.

Compared with the classical minimal reduction problem, there are several differences as follows. The first is the input, where the test costs and measurement errors are the external information. The second is the optimization objective, which is to minimize the test cost instead of the number of features. We can adopt the addition-deletion strategy [15] to design our heuristic reduction algorithm.

In order to address the constraint of the problem, we have defined an inconsistent object in [4]. Here we redefine it as follows.

Definition 10. Let $S = (U, C, D, V, I, n)$ be a decision system with normal distribution measurement errors, $B \subseteq C$, and $x \in U$. In $n_B(x)$, any $x^* \in n_B(x)$ is called an inconsistent object if $D(x^*) \neq D(x)$. The set of inconsistent objects in $n_B(x)$ is

$$ic_B(x) = \{x^* \in n_B(x) \mid D(x^*) \neq D(x)\}. \quad (17)$$

We can evaluate the characteristics of the neighborhood block through the number of inconsistent objects, namely, $|ic_B(x)|$. From Definition 10 we know that given $B \subseteq C$, $x \in \mathrm{POS}_C(D)$ if and only if $ic_B(x) = \emptyset$. Consequently, the $ic_B(x)$ is an important parameter when we compute the positive region. Therefore, the following proposition can be used as an alternative definition of a reduct.

Proposition 11. *Let $S = (U, C, D, V, I, n)$ be a NEDS. Any $R \subseteq C$ is a decision-relative reduct if and only if*

(1) $\forall x \in \mathrm{POS}_C(D)$, $ic_R(x) = \emptyset$,

(2) $\forall x \in R, \exists x \in \mathrm{POS}_C(D), st.\ ic_{R-\{a\}}(x) \neq \emptyset$.

Sometimes we are interested in minimal reduction or minimal test cost reduct (see, e.g., [6]). In this work, we focus on finding reducts with minimal test cost, that is, test-cost-sensitive attribute reducts. Because the TCI-NEDS is a natural extension of NEDS, concepts in the NEDS are also applicable to the TCI-NEDS. We introduce the following definition.

Definition 12. Let $\mathrm{Red}(S)$ denote the set of all reducts of a TCI-NEDS $S = (U, C, D, V, I, n, c)$. Any $R \in \mathrm{Red}(S)$ where $c(R) = \min\{c(R') \mid R' \in \mathrm{Red}(S)\}$ is called a *minimal test cost reduct*.

According to this definition, we should compute all reducts firstly. Consequently, exhaustive algorithms are needed to address this problem. However, for large datasets, finding reducts with minimal test cost is NP hard. Therefore, we should propose a heuristic algorithm to deal with this problem for large datasets.

3.3. Evaluation Measures. In order to dispel the influence of subjective and objective factors, three evaluation measures are adopted to evaluate the performance of the proposed algorithm. We adopt the three measures proposed in [6] for this purpose. These are finding optimal factor (FOF), maximal exceeding factor (MEF), and average exceeding factor (AEF).

Let K be the number of experiments and let k be the number of searched optimal reduct in the experiments. The finding optimal factor is a both qualitative and quantitative measure, which is defined as

$$op = \frac{k}{K}. \quad (18)$$

The maximal exceeding factor describes the worst case of an algorithm, which is defined as

$$\max_{1 \leq i \leq K} ef(R_i), \quad (19)$$

where $ef(R) = (c(R) - c(R'))/(c(R'))$ is the exceeding factor indicating the badness of a reduct, which is a quantitative measure, where R' is an optimal reduct and R is the searched reduct.

The average exceeding factor is defined as

$$\frac{\sum_{i=1}^{K} ef(R_i)}{K}, \quad (20)$$

which represents the whole performance of an algorithm.

4. Algorithm

Test-cost-sensitive attribute reduct problem is NP-hard problem. Therefore, heuristic algorithms are needed to calculate the possible reducts for large datasets. In order to evaluate the performance of a heuristic algorithm, we should find an optimal reduct from all reducts. Hence, exhaustive algorithms are also needed.

In this section, we mainly present a heuristic algorithm and a competition approach to deal with the new problem. The exhaustive algorithm of [4] is adopted to find all reducts of datasets. It is based on backtracking where pruning techniques are crucial in reducing computation.

4.1. The δ-Weighted Heuristic Reduction Algorithm. To design a heuristic algorithm, we employ an algorithm framework

Input: $(U, C, D, \{V_a\}, \{I_a\}, n, c)$
Output: A reduct with minimal test cost
Method:
(1) $B = \emptyset$;
//Attribute addition
(2) $C_t = C$;
(3) **while** $(\mathrm{POS}_B(D) \neq \mathrm{POS}_C(D))$ **do**
(4) **for** each $a \in C_t$ **do**
(5) Compute $f(B, a, c)$;
(6) **end for**
(7) Select a^* with the maximal $f(B, a^*, c)$;
(8) $B = B \cup \{a^*\}$; $C_t = C_t - \{a^*\}$;
(9) **end while**
//Attribute deletion
(10) $C_t = B$;
(11) **while** $(C_t \neq \emptyset)$ do
(12) Select a^* with the maximal test cost.
(13) $C_t = C_t - \{a^*\}$;
(14) **if** $(\mathrm{POS}_{B-\{a^*\}}(D) = \mathrm{POS}_B(D))$ **then**
(15) $B = B - \{a^*\}$;
(16) **end if**
(17) **end while**
(18) return B;

Algorithm 1: An addition-deletion test-cost-sensitive reduction algorithm.

which is similar to the one proposed in [6]. The algorithm follows the typical addition-deletion strategies [15], which is listed in Algorithm 1. It constructs a superreduct and then reduces it to obtain a reduct. The algorithm is essentially different from the one in [6]. First, the input S is a test-cost-independent decision system with normal distribution measurement errors, which is more generalization than the TCI-ER. Second, test results are numerical with normal distribution measurement errors rather than only nominal. The key code of this framework is listed in lines 5 and 7, and the attribute significance function is redefined to obtain respective algorithm. The efficiency of the δ-weighted heuristic reduction algorithm will be discussed in Section 5.4.

As previously mentioned, $|ic_B(x)|$ is an important parameter in evaluating the quality of a neighborhood block. Now, we introduce the following concepts.

Definition 13. Let $S = (U, C, D, V, I, n)$ be a NEDS, $B \subseteq C$, and $x \in U$. $|ic_B(x)|$ is the number of inconsistent objects in neighborhood $n_B(x)$. The total number of inconsistent objects with respect to the positive region is

$$pc_B(S) = \sum_{x \in \mathrm{POS}_C(D)} \left| ic_B(x) \right|. \tag{21}$$

Finally, we propose a δ-weighted heuristic information function:

$$f\left(B, a_i, c\left(a_i\right)\right) = \Phi + \delta \frac{\Phi}{c\left(a_i\right)}, \tag{22}$$

where $\Phi = pc_B(S) - pc_{B \cup \{a_i\}}(S)$ is necessary and indispensable, and it plays a dominant role in the heuristic information, where $c(a_i)$ is the test cost of a_i and $\delta \geq 0$ is a user-specified parameter. If $\delta = 0$, test costs are essentially not considered. If $\delta > 0$, tests with lower cost have bigger significance. Different δ settings can adjust the significance of test cost.

Table 2: Database information.

No.	Name	Domain	$\lvert U\rvert$	$\lvert C\rvert$	$\lvert C'\rvert$	$D = d$
1	Iris	Zoology	150	4	4	Class
2	Glass	Manufacture	214	9	13	Type
3	Wine	Agriculture	178	13	21	Class
4	Wpbc	Clinic	198	33	65	Outcome
5	Wdbc	Clinic	569	30	58	Diagnosis
6	Credit	Commerce	690	15	23	Class
7	Image	Graphics	210	19	30	Class
8	Iono	Physics	351	34	68	Class
9	Liver	Clinic	345	6	8	Selector
10	Diab	Clinic	768	8	12	Class

4.2. The Competition Approach. In order to obtain better results, the competition approach has been discussed in [6]. In the new environment, it is still valid because there is no universally optimal δ. In this approach, reducts complete against each other with only one winner, that is, a reduct with minimal test cost, which can be obtained using $\delta \in L$:

$$c_L = \min_{\delta \in L} c\left(R_\delta\right), \tag{23}$$

where R_δ is the reduct obtained by Algorithm 1 using the heuristic information, with the $|L|$ sets of user-specified δ values.

This approach can improve the quality of the result significantly. The algorithm runs $|L|$ times with different δ values; hence, more runtime is needed. However, it is acceptable because the heuristic algorithm is fast.

5. Experiments

5.1. Data Generation. Most datasets from the UCI library [53] have no intrinsic measurement errors and test costs. Therefore, in our experiments, we create these data to study the performance of the reduction algorithm. For example, measurement errors satisfy normal distribution and Pauta criterion. For the same object, the condition attribute with narrower error ranges should be more expensive. In this section, we will discuss the generation of both the measurement errors and test costs.

Step 1. Ten datasets from the UCI Repository of Machine Learning Databases are selected, and these datasets are listed in Table 2. Each dataset should contain exactly one decision attribute and have no missing value. In order to facilitate processing, firstly, we normalize the data items from their value into a range from 0 to 1. And then, missing values are directly set to 0.5.

Step 2. We produce the number of additional tests for one particular attribute. We generate a random integer k^* in the

Table 3: Generated error ranges for different databases.

Datasets	Minimal	Maximal	Average
Iris	0.0042	0.0044	0.0043
Glass	0.0005	0.0059	0.0030
Wine	0.0031	0.0053	0.0040
Wpbc	0.0011	0.0056	0.0031
Wdbc	0.0006	0.0040	0.0023
Credit	0.0001	0.0056	0.0022
Image	0.0001	0.0065	0.0026
Iono	0.0045	0.0087	0.0061
Liver	0.0011	0.0065	0.0029
Diab	0.0009	0.0059	0.0031

range $[0, k]$ and $k \leq 5$. That is, we have $(k^* + 1)$ measurement methods to obtain values for each object. The number of tests including the additional ones for our experiments is $|C'|$, which is listed in Table 2.

Step 3. We produce the $n(a)$ for each original test according to (6) and (7). The $n(a)$ is computed according to the value of databases without any subjectivity. Three kinds of error ranges of different databases are shown in Table 3. These error ranges are maximal, minimal, and average error ranges of all attributes, respectively. The precision of $n(a)$ can be adjusted through Δ setting, and we set Δ to be 0.01 in our experiments.

Step 4. We produce "new" data subject to error ranges. Letting a_1 be the original test, we can add a random number in $[-(i-1)n(a), (i-1)n(a)]$ to $a_1(x)$ to produce $a_i(x)$, where $x \in U$. The number is generated as follows.

Letting x_1 and x_2 be uniformly distributed on $(0, 1)$, then

$$y_1(x_1, x_2) = \sqrt{-2 \ln x_1} \cos(2\pi x_2) \tag{24}$$

is a random number which has a normal distribution with $\mu = 0$ mean and $\sigma^2 = 1$. From Proposition 4 we know the $n(a) = 3\sigma$, and $\sigma = (1/3)n(a)$.

Since we need a random number in $[-n(a), +n(a)]$, we let

$$y(n(a), x_1, x_2) = \frac{1}{3} y_1(x_1, x_2)\, n(a). \tag{25}$$

Finally, the error range is

$$\text{NDME} = \begin{cases} -n(a) & \text{if } y < -n(a), \\ n(a) & \text{if } y > n(a), \\ y(n(a), x_1, x_2) & \text{otherwise.} \end{cases} \tag{26}$$

According to Definition 8, $\pm i * n(a)$ is the error range of the new test a_i.

The generated NDME with different error ranges are shown in Figure 4. The generated NDME of different databases are shown in Figure 5.

Step 5. The test costs are produced, and they are always represented by positive integers. Let a_1 be the original test and let a_1 be the last test for one particular data item. $c(a_l)$ is set to a random number in $[1, 100]$. $c(a_i)$ where $1 \leq i < l$ is set to $2 \times c(a_{i+1})$. This setting guarantees that tests with narrower error ranges are more expensive.

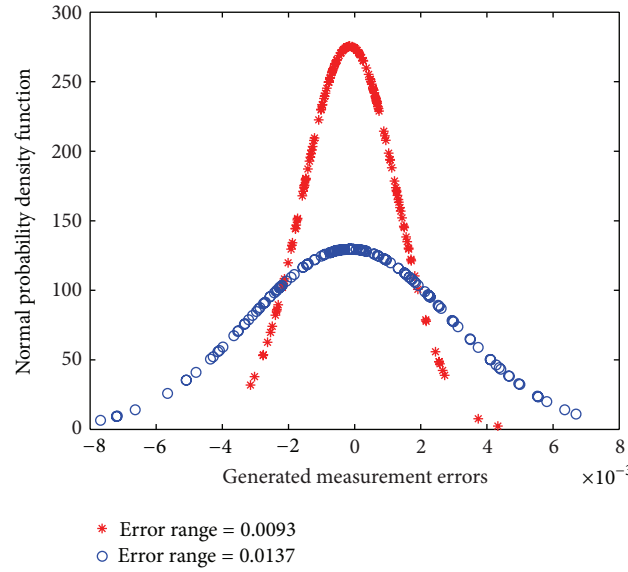

Figure 4: Normal distribution measurement errors with different error ranges.

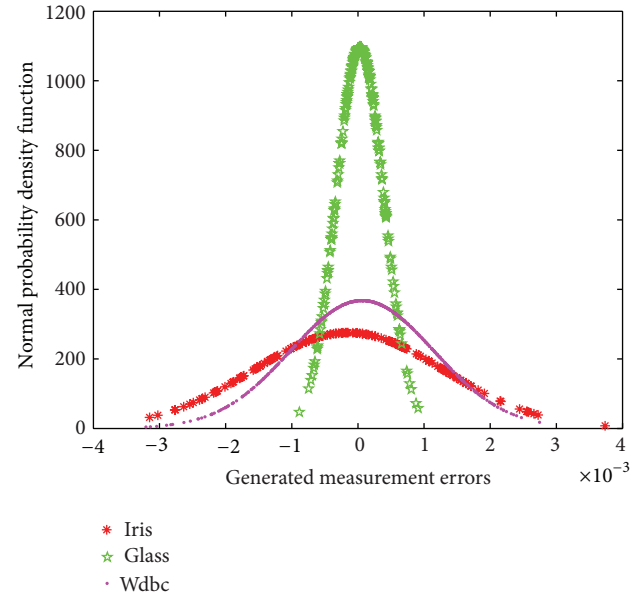

Figure 5: Normal distribution measurement errors of Iris, Glass, and Wdbc.

A dataset generated by this approach is listed in Table 4. SL stands for sepal length, SW stands for sepal width, PL stands for petal length, and PW stands for petal width. SL-1, PL-1, and PL-2 indicate different revisions of the original data. There is only one method to obtain SW and PW.

5.2. Effectiveness of the Heuristic Algorithm. Let $\delta = 2, 3, 4, \ldots, 9$. The heuristic algorithm runs 800 times with different test cost settings and different δ setting on all datasets. Figures 6 and 7 show the results of FOF. For different

Table 4: A generated measurement error vector and a generated test cost vector (Iris).

a	SL	SL-1	SW	PL	PL-1	PL-2	PW
Original test	True	False	True	True	False	False	True
$e(a)$	0.0043	0.0086	0.0041	0.0041	0.0082	0.0123	0.0041
$c(a)$	28	14	81	376	188	94	91

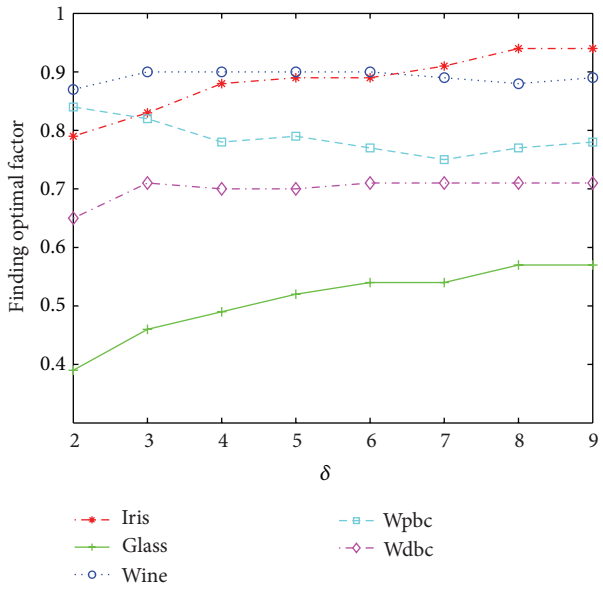

Figure 6: Finding optimal factor (datasets 1–5).

settings of δ, the performance of the algorithm is completely different; that is, the test cost plays a key role in this heuristic algorithm. The results are incomparable to others when $\delta = 0$; hence, these results are not included in this experiment results.

Figures 8 and 9 show the results of MEF, which provide the worst case of the algorithm, and they should be viewed as a statistical measure. Figures 10 and 11 show the results of AEF. These display the overall performance of the algorithm from a statistical perspective.

From the results we observe that the quality of the results varies for different datasets. It is related to the dataset itself because the error range and heuristic information are all computed according to the values of dataset. Then the AEF is less than 0.3 in most cases, which is acceptable. Although the results are generally acceptable, the performance of the algorithm should be improved. In Section 5.3, we will address this issue further.

5.3. Comparison of Three Approaches. For the proposed algorithm, δ is a key parameter. Three approaches can be obtained with different setting of δ. The first approach is called nonweighting approach. It is the only one without taking into account test costs, which is implemented by setting $\delta = 0$. The second approach, called the best δ approach, is to choose the best δ value as depicted in Figures 6 through 11. The third approach is the competition

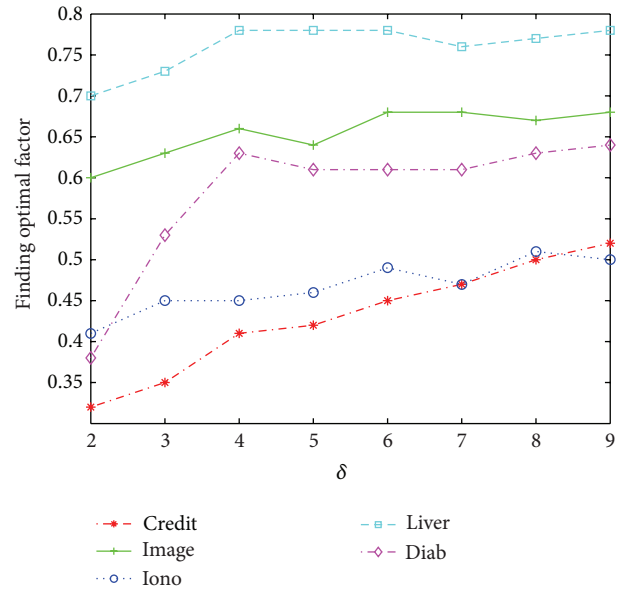

Figure 7: Finding optimal factor (datasets 6–10).

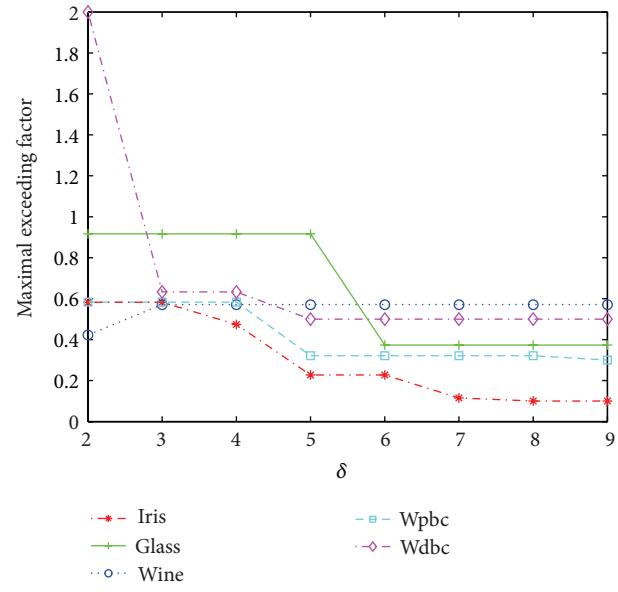

Figure 8: Maximal exceeding factor (datasets 1–5).

approach discussed in Section 4.2. All three are based on Algorithm 1 and the same databases. Now we compare the performance of the proposed algorithm through three approaches mentioned in Section 4.

Table 5 lists results for all three approaches. From Table 5, we observe the following.

(1) The nonweighting approach almost does not find the optimal reduct. It is unacceptable from all three metrics.

(2) In most cases, the best δ approach obtains optimal results. However, we have no idea how to obtain the best value of δ in real applications.

Table 5: Results for $\delta = 0, \delta$ with the optimal setting, and δ with a number of choices.

Dataset	FOF			MEF			AEF		
	$\delta = 0$	$\delta = \delta^*$	$\delta \in L$	$\delta = 0$	$\delta = \delta^*$	$\delta \in L$	$\delta = 0$	$\delta = \delta^*$	$\delta \in L$
Iris	0.170	0.940	0.940	2.000	0.100	0.100	0.360	0.003	0.003
Glass	0.090	0.570	0.640	3.220	0.374	0.374	0.700	0.064	0.049
Wine	0.000	0.900	0.940	19.44	0.423	0.423	4.464	0.021	0.014
Wpbc	0.000	0.840	0.880	45.67	0.300	0.250	14.50	0.033	0.017
Wdbc	0.000	0.710	0.760	93.20	0.500	0.500	14.61	0.041	0.037
Credit	0.000	0.520	0.550	2.188	0.317	0.310	1.095	0.053	0.049
Image	0.000	0.680	0.790	31.43	0.406	0.269	5.417	0.053	0.032
Iono	0.000	0.500	0.630	46.60	0.765	0.544	10.28	0.084	0.054
Liver	0.040	0.780	0.910	4.125	0.275	0.181	0.921	0.023	0.008
Diab	0.000	0.640	0.700	3.788	0.481	0.481	1.278	0.048	0.033

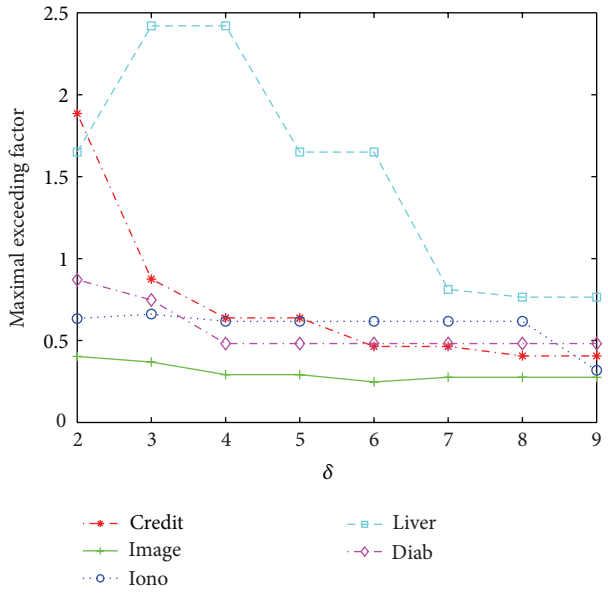

Figure 9: Maximal exceeding factor (datasets 6–10).

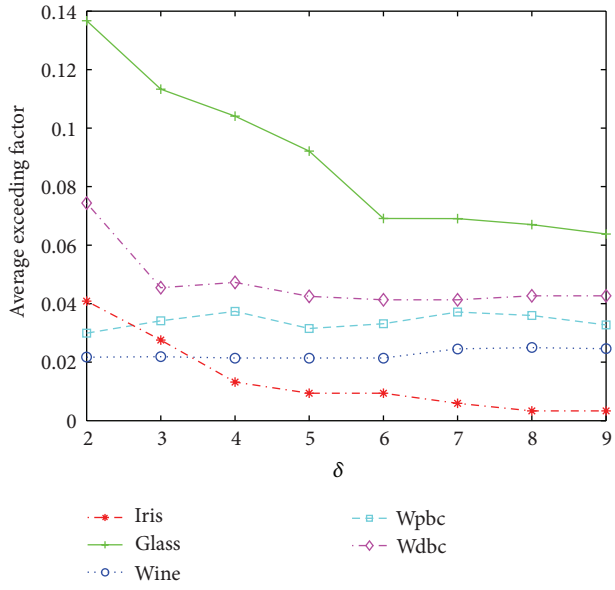

Figure 10: Average exceeding factor (datasets 1–5).

(3) The competition approach improves the quality of results significantly, and the runtime is acceptable for relatively small number of δ.

5.4. Comparison with Existing Algorithm. Compared with an existing model [4], the major improvement is introduced in this section.

First, the NDME was considered to data model, and covering-based rough set based on NDME has been proposed. In most cases, the measurement errors satisfy normal distribution instead of uniform distribution; hence, this new model has wider application areas.

Second, comparing with the fix error range of different databases from [4], the proposed error ranges are adaptively generated according to the database values. Table 3 shows the generated error ranges for different databases. The error ranges for different attributes of the same database are completely different. For example, the maximal error range of Wdbc is 0.0040, and the minimal one is 0.0006.

Third, a δ-weighted heuristic algorithm is developed to deal with the minimal test cost reduct problem. Our algorithm is compared with the λ-weighted algorithm [4] from effectiveness and efficiency. Since two different algorithms have different parameters, we compare the results of the competition approach on ten datasets. Figure 12 shows competition approach results of two algorithms. From the results we observe that

(1) on Wpbc and Iono datasets, two algorithms have the same performance;

(2) λ-weighted algorithm has better performance than our algorithm on Iris, Glass, and Credit datasets;

(3) however, our algorithm performs better than the λ-weighted algorithm on five datasets.

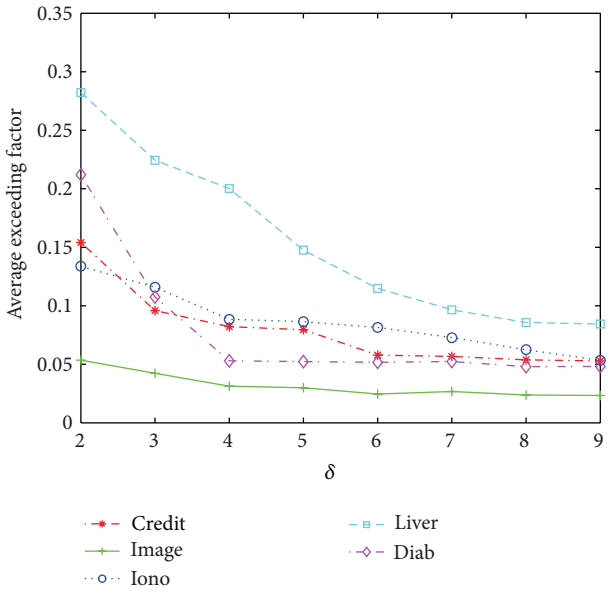

Figure 11: Average exceeding factor (datasets 6–10).

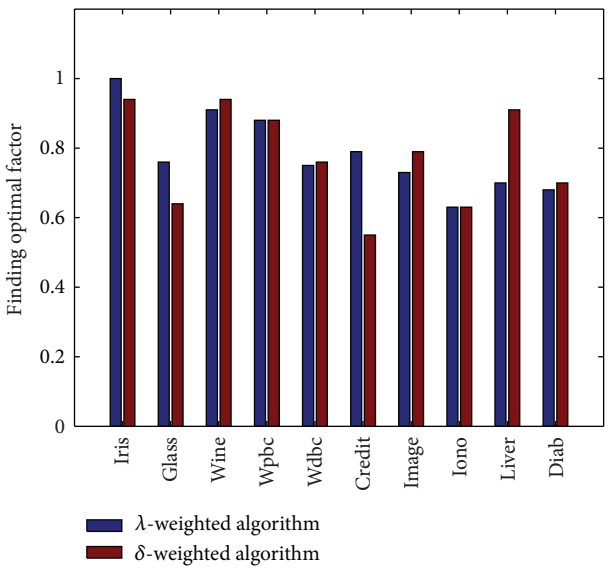

Figure 12: Competition approach results of two algorithms.

The efficiency comparison between the δ-weighted algorithm and λ-weighted one is depicted in Figure 13. From the results we note that our algorithm has an improvement in terms of runtime. Figure 14 shows the efficiency ratios of the δ-weighted algorithm and the λ-weighted algorithm.

6. Conclusions

In rough set model, measurement errors and test costs are all intrinsic to data. In this paper, we built a new covering-based rough set model considering measurement errors and test costs at four levels.

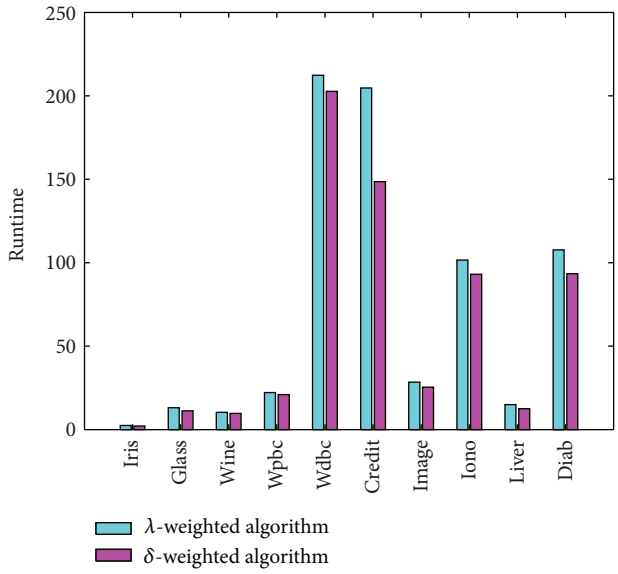

Figure 13: Efficiency comparison.

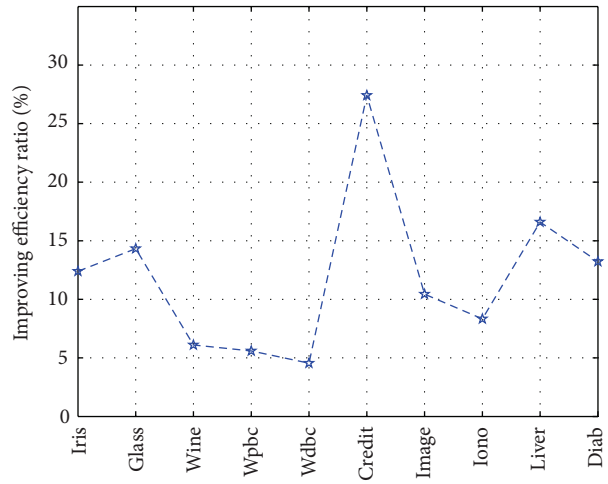

Figure 14: Improving efficiency ratio.

(1) At the data model level, a new data model with NDME and test cost was proposed. This model has more application areas because the measurement errors have certain universality.

(2) At the computational model level, we introduced a covering-based rough set with NDME. This model is generally more complex than that presented in this field.

(3) At the problem level, a minimal test cost reduct problem based on the new model was redefined.

(4) At the algorithm level, a δ-weighted heuristic algorithm was developed to deal with this reduct problem. Experimental results indicate the effectiveness and efficiency of the algorithm.

In summary, the data model based on normal distribution measurement errors has the wide application scope. This

study suggests new research trends of covering-based rough set and cost-sensitive learning.

Acknowledgments

This work is in part supported by National Science Foundation of China under Grant no. 61170128, the Natural Science Foundation of Fujian Province, China, under Grant no. 2012J01294, State Key Laboratory of Management and Control for Complex Systems Open Project under Grant no. 20110106, and Fujian Province Foundation of Higher Education under Grant no. JK2012028, and the Education Department of Fujian Province under Grant no. JA12222.

References

[1] S. Bell, *A Beginner's Guide to Uncertainty of Measurement*, National Physical Laboratory, 2001.

[2] C. C. Aggarwal, "On density based transforms for uncertain data mining," in *Proceedings of IEEE 23rd International Conference on Data Engineering (ICDE '07)*, pp. 866–875, April 2007.

[3] M. Chau, R. Cheng, B. Kao, and J. Ng, "Uncertain data mining: an example in clustering location data," in *Proceedings of the 10th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD '06)*, vol. 3918 of *Lecture Notes in Computer Science*, pp. 199–204, 2006.

[4] F. Min and W. Zhu, "Attribute reduction of data with error ranges and test costs," *Information Sciences*, vol. 211, pp. 48–67, 2012.

[5] W. Fan, S. Stolfo, J. Zhang, and P. Chan, "Adacost: misclassification cost-sensitive boosting," in *Proceedings of the 16th International Conference on Machine Learning (ICML '99)*, 1999.

[6] F. Min, H. P. He, Y. H. Qian, and W. Zhu, "Test-cost-sensitive attribute reduction," *Information Sciences*, vol. 181, pp. 4928–4942, 2011.

[7] F. Min and Q. Liu, "A hierarchical model for test-cost-sensitive decision systems," *Information Sciences*, vol. 179, no. 14, pp. 2442–2452, 2009.

[8] M. Pazzani, C. Merz, P. M. K. Ali, T. Hume, and C. Brunk, "Reducing misclassification costs," in *Proceedings of the 11th International Conference of Machine Learning (ICML' 94)*, Morgan Kaufmann, 1994.

[9] H. Zhao, F. Min, and W. Zhu, "Test-cost-sensitive attribute reduction based on neighborhood rough set," in *Proceedings of the IEEE International Conference on Granular Computing*, 2011.

[10] Z. H. Zhou and X. Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.

[11] Y. Y. Yao and Y. Zhao, "Attribute reduction in decision-theoretic rough set models," *Information Sciences*, vol. 178, no. 17, pp. 3356–3373, 2008.

[12] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artificial Intelligence*, vol. 151, no. 1-2, pp. 155–176, 2003.

[13] X. Y. Jia, W. H. Liao, Z. M. Tang, and L. Shang, "Minimum cost attribute reduction in decision-theoretic rough set models," *Information Sciences*, vol. 219, pp. 151–167, 2013.

[14] H. X. Li, X. Z. Zhou, J. B. Zhao, and D. Liu, "Attribute reduction in decision-theoretic rough set model: a further investigation," in *Proceedings of Rough Sets and Knowledge Technology*, vol. 6954 of *Lecture Notes in Computer Science*, 2011.

[15] Y. Y. Yao, Y. Zhao, and J. Wang, "On reduct construction algorithms," in *Proceedings of the Rough Sets and Knowledge Technology (RSKT '06)*, vol. 4062, pp. 297–304, 2006.

[16] W. Zhu and F.-Y. Wang, "Reduction and axiomization of covering generalized rough sets," *Information Sciences*, vol. 152, pp. 217–230, 2003.

[17] L. W. Ma, "On some types of neighborhood-related covering rough sets," *International Journal of Approximate Reasoning*, vol. 53, no. 6, pp. 901–911, 2012.

[18] W. Zhu, "Generalized rough sets based on relations," *Information Sciences*, vol. 177, no. 22, pp. 4997–5011, 2007.

[19] W. Zhu, "Topological approaches to covering rough sets," *Information Sciences*, vol. 177, no. 6, pp. 1499–1508, 2007.

[20] W. Zhu, "Relationship among basic concepts in covering-based rough sets," *Information Sciences*, vol. 179, no. 14, pp. 2478–2486, 2009.

[21] W. Zhu, "Relationship between generalized rough sets based on binary relation and covering," *Information Sciences*, vol. 179, no. 3, pp. 210–225, 2009.

[22] W. Zhu and F. Wang, "Covering based granular computing for conflict analysis," in *Intelligence and Security Informatics*, vol. 3975 of *Lecture Notes in Computer Science*, pp. 566–571, 2006.

[23] W. Zhu and F. Y. Wang, "On three types of covering-based rough sets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 8, pp. 1131–1143, 2007.

[24] Q. Hu, W. Pedrycz, D. R. Yu, and J. Lang, "Selecting discrete and continuous features based on neighborhood decision error minimization," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 40, no. 1, pp. 137–150, 2010.

[25] Q. H. Hu, D. R. Yu, J. F. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, 2008.

[26] Q. H. Hu, D. R. Yu, and Z. X. Xie, "Numerical attribute reduction based on neighborhood granulation and rough approximation," *Journal of Software*, vol. 19, no. 3, pp. 640–649, 2008 (Chinese).

[27] H. X. Li, M. H. Wang, X. Z. Zhou, and J. B. Zhao, "An interval set model for learning rules from incomplete information table," *International Journal of Approximate Reasoning*, vol. 53, no. 1, pp. 24–37, 2012.

[28] W. Wei, J. Liang, and Y. Qian, "A comparative study of rough sets for hybrid data," *Information Sciences*, vol. 190, pp. 1–16, 2012.

[29] F. Min, W. Zhu, H. Zhao, G. Y. Pan, J. B. Liu, and Z. L. Xu, "Coser: Cost-senstive rough sets," 2012, http://grc.fjzs.edu.cn/~fmin/coser/.

[30] Wikipedia, http://www.wikipedia.org/.

[31] H. Zhao, F. Min, and W. Zhu, "Inducing covering rough sets from error distribution," *Journal of Information & Computational Science*, vol. 10, no. 3, pp. 851–859, 2013.

[32] P. Zhu, "Covering rough sets based on neighborhoods: an approach without using neighborhoods," *International Journal of Approximate Reasoning*, vol. 52, no. 3, pp. 461–472, 2011.

[33] Y. Chen, D. Miao, and R. Wang, "A rough set approach to feature selection based on ant colony optimization," *Pattern Recognition Letters*, vol. 31, no. 3, pp. 226–233, 2010.

[34] Q. Hu, S. An, and D. Yu, "Soft fuzzy rough sets for robust feature evaluation and selection," *Information Sciences*, vol. 180, no. 22, pp. 4384–4400, 2010.

[35] Y. Du, Q. Hu, P. F. Zhu, and P. J. Ma, "Rule learning for classification based on neighborhood covering reduction," *Information Sciences*, vol. 181, no. 24, pp. 5457–5467, 2011.

[36] Y. H. Qian, J. Y. Liang, and C. Y. Dang, "Converse approximation and rule extraction from decision tables in rough set theory," *Computers & Mathematics with Applications*, vol. 55, no. 8, pp. 1754–1765, 2008.

[37] X. B. Yang, D. J. Yu, J. Y. Yang, and X. N. Song, "Difference relation-based rough set and negative rules in incomplete information system," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 17, no. 5, pp. 649–665, 2009.

[38] Z. Pawalk, "Rough sets: theoretical aspects of reasoning about data," 1991.

[39] Z. Pawlak and A. Skowron, "Rough sets: some extensions," *Information Sciences*, vol. 177, no. 1, pp. 28–40, 2007.

[40] W. Z. Wu and Y. Leung, "Theory and applications of granular labelled partitions in multi-scale decision tables," *Information Sciences*, vol. 181, no. 18, pp. 3878–3897, 2011.

[41] Y. Qian, J. Liang, D. Li, H. Zhang, and C. Dang, "Measures for evaluating the decision performance of a decision table in rough set theory," *Information Sciences*, vol. 178, no. 1, pp. 181–202, 2008.

[42] R. B. Barot and T. Y. Lin, "Granular computing on covering from the aspects of knowledge theory," in *Proceedings of the IEEE Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS '08)*, May 2008.

[43] S. Calegari and D. Ciucci, "Granular computing applied to ontologies," *International Journal of Approximate Reasoning*, vol. 51, no. 4, pp. 391–409, 2010.

[44] T. Y. Lin, "Granular computing: practices, theories, and future directions," *Encyclopedia of Complexity and Systems Science*, vol. 2009, pp. 4339–4355, 2009.

[45] L. Zadeh, "Fuzzy sets and information granularity," *Advances in Fuzzy Set Theory and Applications*, vol. 11, pp. 3–18, 1979.

[46] A. Bargiela and W. Pedrycz, *Granular Computing: An Introduction*, Kluwer Academic, Boston, Mass, USA, 2003.

[47] T. Y. Lin, "Granular computing on binary relations-analysis of conflict and chinese wall security policy," in *Proceedings of the Rough Sets and Current Trends in Computing*, vol. 2475 of *Lecture Notes in Computer Science*, 2002.

[48] T. Y. Lin, "Granular computing structures, representations, and applications," in *Proceedings of the 9th International Conference (RSFDGrC '03)*, vol. 2639 of *Lecture Notes in Artificial Intelligence*, pp. 16–24, May 2003.

[49] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[50] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic, Boston, Mass, USA, 1991.

[51] J. G. Bazan and A. Skowron, "Dynamic reducts as a tool for extracting laws from decision tables," in *Proceedings of the 8th International Symposium on Methodologies for Intelligent Systems*, 1994.

[52] Y. H. Qian, J. Y. Liang, W. Pedrycz, and C. Y. Dang, "Positive approximation: an accelerator for attribute reduction in rough set theory," *Artificial Intelligence*, vol. 174, no. 9-10, pp. 597–618, 2010.

[53] C. L. Blake and C. J. Merz, "UCI repository of machine learning databases," 1998, http://www.ics.uci.edu/~mlearn/mlrepository.html.