

## Research Article

# A Novel Approach for Reducing Attributes and Its Application to Small Enterprise Financing Ability Evaluation

Baofeng Shi <sup>1</sup>, Bin Meng <sup>2</sup>, Hufeng Yang,<sup>1</sup> Jing Wang,<sup>1</sup> and Wenli Shi <sup>3</sup>

<sup>1</sup>College of Economics & Management, Northwest A&F University, Yangling, Shaanxi 712100, China

<sup>2</sup>Collaborative Innovation Center for Transport Studies, Dalian Maritime University, Dalian, Liaoning 116026, China

<sup>3</sup>School of Management Engineering and Business, Hebei University of Engineering, Handan, Hebei 056038, China

Correspondence should be addressed to Wenli Shi; shiwenli2008@126.com

Received 28 June 2017; Accepted 5 December 2017; Published 15 January 2018

Academic Editor: Dimitri Volchenkov

Copyright © 2018 Baofeng Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Attribute reduction is viewed as a kind of preprocessing steps for reducing large dimensionality in data mining of all complex systems. A great deal of researchers have proposed various approaches to reduce attributes or select key features in multicriteria decision making evaluation. In practice, the existing approaches for attribute reduction focused on improving the classification accuracy or saving the cost of computational time, without considering the influence of the reduction results on the original data set. To help address this gap, we develop an advanced novel attribute reduction approach combining Pearson correlation analysis with  $F$  test significance discrimination for the screening and identification of key characteristics related to the original data set. The proposed model has been verified using the financing ability evaluation data of 713 small enterprises of a city commercial bank in China. And the experimental results show that the proposed reduction model is efficient and effective. Moreover, our experimental findings help to locate the qualified partners and alleviate the difficulties faced by enterprises when applying loan.

## 1. Introduction

With the coming of the era of big data, the size of data sets has been increasing sharply, causing the decision makers and management to have difficulty in making decisions based on those data [1]. Then the most important thing for decision makers is to reduce huge attributes or large dimensionality in data sets. Attribute reduction, also called indicators selection or feature screening, ascertains a subset of attributes to reduce the dimensionality of the original data sets. Utilizing reducing attributes, it can select the attributes with the highest information content and save the cost of computational time and memory [2]. Besides, it is also useful to improve the classification accuracy as a result of deleting the information chaos and irrelevant attributes [3]. In practice, attribute reduction has been applied to a great deal of fields such as decision making, pattern recognition, and economic and social system evaluation [4–7].

The main attribute reduction approaches can be divided into three categories. One of the most famous methods for

attribute reduction is based on rough set theory. Rough set approach proposed by Pawlak provides useful tools for reasoning from data [8]. It is advantageous to other approaches for attribute reduction that typically use multivariate statistics which require specific parametric assumptions [9, 10]. Degang et al. established a model to reduce the attributes of covering decision systems combining traditional rough set. Empirical study indicated that the proposed attribute reduction approach accomplished better classification performance than those of existing rough set methods [11]. In order to improve the classification accuracy containing hybrid type attributes, such as discretizing numerical attributes or categorical attributes, Hu et al. introduced a simple and efficient greedy algorithm for hybrid attribute reduction [12]. When some decision or evaluation systems have some errors, missing data, and missing attributes in observation, neither DRSA (dominance-based rough set approach) [13] nor VC-DRSA (variable-consistency dominance-based rough set approach) [14] can work appropriately. Inuiguchi et al. created a variable-precision dominance-based rough

set approach (VP-DRSA) to deal with these problems [15]. Tsang et al. presented an attribute reduction model with covering rough sets based on discernibility matrix to compute all attribute reducts [16]. Furthermore, Wang et al. developed a novel approach for constructing simpler discernibility matrix with covering rough sets, and it improved some characterizations of attribute reduction proposed by Wang et al. [17]. In addition, there are the two most important attribute reduction models, which extend the Pawlak's rough set, the neighborhood rough set (NRS) model [18] and the fuzzy rough set model [19]. They can tackle continuous numeric data and fuzzy information granulation, and the determination of what objects should be included in a rough set allowed some flexibility [20].

The second method for screening key factors is the attribute reduction models based on statistics or econometrics technique. In order to obtain preference information of the decision maker in multiobjective search, Zitzler and Künzli defined an optimization goal in terms of a binary performance measure, to select key information directly utilizing this measure [21]. Polat and Krmac screened the most important attributes using pairwise Fisher score attribute reduction approach (PFSAR) and correlation based attribute reduction [22]. Ju and Sohn developed a technology attribute reduction model that uses logistic regression based on exploratory factor analysis (EFA) of 16 technology-related attributes [23]. Elliott et al. developed a model based on a double hidden Markov model (DHMM), to extract information about the "true" credit qualities of firms [24]. Shi et al. created an indicators extraction model based on Pearson correlation analysis and logistic regression significant discriminant in customers' classification. The proposed approach ensured the reserved indicators can effectively distinguish default customers from nondefault customers [25].

In addition, there are other attributes reduction methods, such as the concept lattice model, the heuristic algorithm, and the colony optimization algorithm. Some researchers developed some new attribute reduction models by using the concept lattice classification theory [26–28]. Wei et al. discussed attribute reduction in information systems by establishing three equivalence relations on the attribute set and its power set [29]. In overwhelming data analysis and machine learning studies, most existing attribute reduction work focused on improving the classification accuracy. However, these studies neglected the problem of how to decrease the test cost. Min et al. proposed a heuristic algorithm to handle this problem in attribute reduction [30]. Chi et al. created an indicators screen model based on correlation analysis and component analysis [31]. Minimal test cost attribute reduction is very important in cost-sensitive machine learning. However, in many cases these heuristic algorithms cannot find the optimal solution. In order to deal with this problem, Xu et al. established an ant colony optimization algorithm for attribute reduction. Experimental results on UCI data sets showed that the proposed method outperforms the information gain-based approach [32]. According to the principle of eliminating redundant information and the principle of the maximum information content, Shi and Chi proposed an attribute reduction model combining  $R$  cluster analysis and coefficient of variation [33].

Because people are interested in the maximal rules implicated in attribute reduction, Li et al. developed two new kinds of attribute reduction approaches in the decision formal context based on maximal rules [34].

The existing findings can offer important references for reducing attributes. However, there are still some limitations. First of all, in the evaluation of complex systems, the aim of the attribute reduction is to eliminate the factors, which should not have significant effect on the comprehensive evaluation results. However, the existing attribute reduction approaches have not established the comprehensive index  $y$  (i.e., the comprehensive score vector  $y$ ), which can reflect all of the attributes' characteristics. This means that the existing attribute reduction approaches have not developed the relationship between attributes and the comprehensive index  $y$  (i.e., the comprehensive evaluation result). This results in some reserved attributes, which do not have significant effect on the comprehensive evaluation result. And secondly, most of existing attributes reduction approaches judged the performance of the proposed approach by the standard of saving the cost of computational time. The standard does not analyze the information contribution degree of the reserved attributes to the mass-election attributes. Thirdly, most of existing researches verify the applicability of the proposed attribute reduction methods using numerical simulation, but not utilizing actual data.

To solve the shortcomings, this study creates a novel attribute reduction model to screen the key influencing factors. We advance in three aspects. First, this paper establishes an attribute reduction approach by combining Pearson correlation analysis with  $F$  test significance discrimination. Pearson correlation analysis is applied to the calculation of the correlation among attributes to delete the similar attributes.  $F$  test significance discrimination is used to select the key attributes which have the greatest influence on comprehensive index  $y$ . Second, we also define an information contribution ratio to assess this attribute reduction approach from a statistical viewpoint. Third, the proposed attribute reduction approach has been verified by utilizing the financing ability evaluation data of 713 small enterprises of a city commercial bank in China. Empirical evidence presents that the selected attributes reflect 94.7% original information with 27.54% original attributes. Furthermore, this paper also selects 19 key influencing factors for assessing the financing ability of small enterprises.

The remainder of this paper is organized as follows. Section 2 introduces the design and methodology of this study. Section 3 presents the data and empirical analysis of our attribute reduction model for 713 small enterprises. Section 4 concludes and highlights the future research directions of this paper.

## 2. Design and Methodology of the Study

In this section, we introduce a novel attribute reduction model by combining Pearson correlation analysis with  $F$  test significance discrimination approach. First of all, in order to eliminate the influence of the differences of attributes units and dimensions on attribute reduction, the original data

should be transformed into real numbers within the interval  $[0, 1]$ . Secondly, we utilize Pearson correlation analysis to delete the attributes of large correlation from the whole mass-election attributes set, avoiding repeated information. Thirdly,  $F$  test significance discrimination approach has been created to select the attributes with the highest information content, which ensures that the selected attribute has the greatest influence on the small enterprise financing performance. A step-by-step instruction is as follows.

**2.1. Standardization of Attribute Data.** In our attribute reduction model, the first step is standardization of attribute data so that the after-calculation processes and parameters use the same standard. According to the features of attributes,

$$p_{ij} = \frac{v_{ij} - \min_{1 \leq j \leq n}(v_{ij})}{\max_{1 \leq j \leq n}(v_{ij}) - \min_{1 \leq j \leq n}(v_{ij})} \quad (1)$$

$$p_{ij} = \frac{\max_{1 \leq j \leq n}(v_{ij}) - v_{ij}}{\max_{1 \leq j \leq n}(v_{ij}) - \min_{1 \leq j \leq n}(v_{ij})} \quad (2)$$

$$p_{ij} = \begin{cases} 1 - \frac{q_1 - v_{ij}}{\max(q_1 - \min_{1 \leq j \leq n}(v_{ij}), \max_{1 \leq j \leq n}(v_{ij}) - q_2)}, & v_{ij} < q_1 \quad (a) \\ 1 - \frac{v_{ij} - q_2}{\max(q_1 - \min_{1 \leq j \leq n}(v_{ij}), \max_{1 \leq j \leq n}(v_{ij}) - q_2)}, & v_{ij} > q_2 \quad (b) \\ 1, & q_1 \leq v_{ij} \leq q_2 \quad (c) \end{cases} \quad (3)$$

where  $p_{ij}$  is the standardized score of the  $j$ th small enterprise on the  $i$ th attribute,  $v_{ij}$  is the attribute original data of the  $j$ th small enterprise on the  $i$ th attribute,  $n$  is the number of small enterprises,  $q_1$  is the left boundary of the ideal interval, and  $q_2$  is the right boundary of the ideal range.

The qualitative attributes refer to these attributes whose attribute values are described by a text, rather than a numerical value. The standard scores of qualitative attributes can be obtained by rational analysis and expert investigation.

**2.2. Pearson Correlation Coefficients.** The Pearson product-momentum correlation coefficient was developed by Karl Pearson from a related idea introduced by Francis Galton in the 1880s [36]. It is a measure of the linear correlation (dependence) between two random variables. It was also called the PPMCC, PCC, or Pearson's  $r_{xy}$ . Historically, it is the first formal measure of correlation and it is still one of the most widely used measures of relationship.

The Pearson correlation coefficient of two attributes  $x$  and  $y$  is defined as the covariance of the two variables divided by the product of their standard deviations. The Pearson correlation coefficient is commonly represented by the letter  $r$  and it can be equivalently defined by [37]

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{k=1}^n (x_k - \bar{x})^2} \sqrt{\sum_{k=1}^n (y_k - \bar{y})^2}} \quad (4)$$

the attributes can be divided into two types: quantitative attributes and qualitative attributes. The quantitative attributes include positive attributes, negative attributes, and interval attributes. The positive attributes are attributes showing that the greater their values are, the better the small enterprise financing capacity is. The negative attributes are attributes showing that the less their values are, the better small enterprise financing capacity is. The interval attributes are attributes reasonable only when the original index data are within certain range.

The standardization equations of positive attributes, negative attributes, and interval attributes are represented by (1), (2), and (3), respectively, [35]:

where  $\bar{x} = (1/n) \sum_{j=1}^n x_j$ ,  $\bar{y} = (1/n) \sum_{j=1}^n y_j$  are the mean of  $x$  and  $y$ , respectively. Equation (4) is applied to the calculation of the correlation between two variables  $x$  and  $y$ . The coefficient  $r_{xy}$  ranges from  $-1$  to  $1$  and it is invariant to linear transformations of either variables. A value of  $1$  indicates a total positive correlation between  $x$  and  $y$ , a value of  $0$  implies no correlation between  $x$  and  $y$ , and a value of  $-1$  indicates a total negative correlation.

Some authors have offered guidelines for the interpretation of the Pearson correlation coefficient [38–41]. If the Pearson correlation coefficient of two attributes is greater than  $0.8$  [40, 41], we can conclude that these attributes are information redundancy. In this situation, we should remove one of attributes. In the opposite situation, if the Pearson correlation coefficient is smaller than  $0.8$ , it indicates that these attributes are not information redundancy and should keep these two attributes.

**2.3. Attribute Reduction Model.** In our attribute reduction model, the third step is to select the key attribute which has the greatest influence on comprehensive index  $y$  and deleting the uncorrelated attributes. In this part, we first calculate the attribute weightings using entropy weight approach. And then, we can obtain the financing ability evaluation score  $y$  (i.e., comprehensive index  $y$ ) for every small enterprise. Subsequently, the multiple determination coefficient  $R_{m-1}^2$  between comprehensive index  $y$  and all of these  $m - 1$

attributes can be obtained, and the multiple determination coefficient  $R_{m-2}^2$  between comprehensive index  $y$  and the remaining  $m - 2$  attributes after removing an attribute  $x_i$  can be calculated. By using  $F$  test significance discrimination, these key attributes which have the greatest influence on small enterprise financing ability evaluation are selected. At the same time, the reduction idea—that is, the bigger the difference  $\Delta R^2$  between the multiple repeated determination coefficient  $R_{m-1}^2$  and the multiple determination coefficient  $R_{m-2}^2$  ( $\Delta R^2 = R_{m-1}^2 - R_{m-2}^2$ ), the more the significance to comprehensive evaluation results—is reflected. Thus, the right time to make up the existing attribute reduction approaches cannot reflect the influence of attributes on the comprehensive index  $y$ , because the attribute reduction process has nothing to do with comprehensive index  $y$ .

### 2.3.1. Weighting Attributes Utilizing Entropy Weight Method.

Let  $f_{ij}$  denote the weight of the  $i$ th attribute in the  $j$ th small enterprise, let  $p_{ij}$  denote the standard score of the  $i$ th attribute in the  $j$ th small enterprise, let  $n$  denote the number of small enterprises, and let  $m$  denote the number of attributes.

The subordinate degree function  $f_{ij}$  of the attribute  $p_{ij}$  is given by

$$f_{ij} = \frac{p_{ij}}{\sum_{j=1}^n p_{ij}}. \quad (5)$$

Then, the entropy  $H_i$  of the  $i$ th attribute can be calculated with

$$H_i = -\frac{1}{\ln n} \sum_{j=1}^n f_{ij} \ln f_{ij}. \quad (6)$$

And then, the entropy weight  $w_i$  of the  $i$ th attribute is [42]

$$w_i = \frac{1 - H_i}{m - \sum_{i=1}^m H_i}. \quad (7)$$

where  $\sum_{i=1}^m w_i = 1$ .

**2.3.2. Reducing Attributes Based on  $F$  Test Significance Discrimination.** After eliminating redundant information in Section 2.2, this section will select the key attributes which have the greatest influence on comprehensive index  $y$  utilizing  $F$  test significance discrimination approach. Now we outline the steps to build an attribute reduction model based on  $F$  test significance discrimination.

*Step 1.* Calculate the comprehensive index  $y$ . Let  $y_j$  denote the comprehensive index or the comprehensive score for the  $j$ th small enterprise financing ability evaluation. We have

$$y_j = \sum_{i=1}^m w_i p_{ij}. \quad (8)$$

The meanings of the rest of variables in (8) are the same as the variables in (1) and (7).

*Step 2.* Calculate Pearson correlation coefficients  $r_i$  ( $i = 1, 2, \dots, m$ ) between attribute  $x_i$  and the comprehensive index

$y$ . We can assume the attributes ranking is  $x_1^*, x_2^*, \dots, x_m^*$  according to the correlation coefficient absolute value  $|r_i|$  in a descending order.

*Step 3.* Calculate the multiple determination coefficient  $R_{m-1}^2$  between comprehensive index  $y$  and the remaining  $m - 1$  attributes  $x_2^*, x_3^*, \dots, x_m^*$  after removing the first attribute  $x_1^*$  with the biggest correlation coefficient absolute value.

Let  $a_0, a_2, a_3, \dots, a_m$  denote the estimated parameters, respectively, let  $x_2^*, x_3^*, \dots, x_m^*$  denote  $m - 1$  attributes, and let  $\varepsilon$  denote the random error term. The regression function is given by

$$y = a_0 + a_2 x_2^* + a_3 x_3^* + \dots + a_m x_m^* + \varepsilon. \quad (9)$$

In (9), the estimated values for parameters  $a_0, a_2, a_3, \dots, a_m$  can be obtained using the least squares regression estimation method. Furthermore, the estimated value vector  $\hat{y}$  of the comprehensive index  $y$  can be calculated. Then, we have [43]

$$R_{m-1}^2 = \frac{\sum_{j=1}^n (\hat{y}_j - \bar{y})^2}{\sum_{j=1}^n (y_j - \bar{y})^2}, \quad (10)$$

where  $\bar{y} = (1/n) \sum_{j=1}^n y_j$  and  $n$  denotes the number of small enterprises.

It should be pointed out that the attribute  $x_1^*$  should be reserved in attribute reduction, because the attribute  $x_1^*$  has the maximum pertinency with the comprehensive evaluation results. It also indicates that the attribute  $x_1^*$  has the biggest impact on small enterprise financing ability evaluation.

*Step 4.* Calculate the multiple determination coefficient  $R_{m-2}^2$  between comprehensive index  $y$  and the remaining  $m - 2$  attributes  $x_3^*, x_4^*, \dots, x_m^*$  after removing the first attribute  $x_1^*$  with the biggest correlation coefficient absolute value and the second attribute  $x_2^*$  with the second biggest correlation coefficient absolute value.

Let  $b_0, b_3, b_4, \dots, b_m$  denote the estimated parameters, respectively, let  $x_3^*, x_4^*, \dots, x_m^*$  denote  $m - 2$  attributes, and let  $\eta$  denote the random error term. The regression function is as follows:

$$y = b_0 + b_3 x_3^* + b_4 x_4^* + \dots + b_m x_m^* + \eta. \quad (11)$$

In the same way, we can calculate the estimated value vector  $\hat{y}'$  of the comprehensive index  $y$  for (11). And the multiple determination coefficient  $R_{m-2}^2$  is given by

$$R_{m-2}^2 = \frac{\sum_{j=1}^n (\hat{y}'_j - \bar{y})^2}{\sum_{j=1}^n (y_j - \bar{y})^2}. \quad (12)$$

*Step 5.* Calculate  $\Delta R^2$ . Let  $\Delta R^2$  denote the difference of the multiple determination coefficient  $R_{m-1}^2$  and the multiple determination coefficient  $R_{m-2}^2$ ; namely,

$$\Delta R^2 = R_{m-1}^2 - R_{m-2}^2. \quad (13)$$



In (13), the difference  $\Delta R^2$  reflects the influence of the attribute  $x_2^*$  on the comprehensive index  $y$ . If  $\Delta R^2$  is not equal to zero significantly, it means that the attribute  $x_2^*$  affects the comprehensive evaluation result  $y$  significantly, and therefore the attribute  $x_2^*$  should be reserved. On the contrary, if  $\Delta R^2$  is equal to zero significantly, then  $\Delta R^2 = 0$ , which indicates the attribute  $x_2^*$  does not have significant effect on the comprehensive evaluation result  $y$ , and the attribute  $x_2^*$  should be deleted.

*Step 6.* Reduce attributes establishing  $F$  test significance discrimination.

Hypothesis  $H_0: \Delta R^2 \neq 0$ ;  $H_1: \Delta R^2 = 0$ .

Let  $F_i$  denote the  $F$  test value of the  $i$ th attribute  $x_i$ ; we have [44]

$$F_i = \frac{\Delta R^2 / 1}{(1 - R_{m-1}^2) / [n - (m - 2)]} \quad (14)$$

$$= \frac{(R_{m-1}^2 - R_{m-2}^2) [n - (m - 2)]}{1 - R_{m-1}^2}.$$

For (14), we can understand its meanings from the following three aspects. Firstly, the bigger the multiple determination coefficient  $R_{m-1}^2$  is, the smaller the deviation of the estimated value  $\hat{y}$  and the actual comprehensive index  $y$  would be. The smaller the multiple determination coefficient  $R_{m-2}^2$  is, the bigger the deviation of the estimated value  $\hat{y}'$  and the actual comprehensive index  $y$  after removing the attribute  $x_2^*$  would be. That is to say, when we remove the attribute  $x_2^*$ , the explanation ability of the  $m-2$  attributes  $x_3^*, x_4^*, \dots, x_m^*$  to the comprehensive evaluation score  $y$  decreases significantly. It also indicates that the attribute  $x_2^*$  has significant effect on the comprehensive evaluation result  $y$  of small enterprises; thus the attribute  $x_2^*$  should be reserved.

Secondly, the bigger the difference  $\Delta R^2$  of the multiple determination coefficient  $R_{m-1}^2$  and the multiple determination coefficient  $R_{m-2}^2$  is, the bigger the difference of the explanation ability  $R_{m-1}^2$  of the  $m-1$  attributes  $x_2^*, x_3^*, \dots, x_m^*$  to the comprehensive evaluation score  $y$  and the explanation ability  $R_{m-2}^2$  of the  $m-2$  attributes  $x_3^*, x_4^*, \dots, x_m^*$  to the comprehensive evaluation score  $y$  would be. It means that the attribute  $x_2^*$  affects the comprehensive evaluation result  $y$  of small enterprises significantly, and the attribute  $x_2^*$  should not be deleted.

Thirdly, the bigger the difference  $\Delta R^2$  (i.e., the bigger the difference value  $R_{m-1}^2 - R_{m-2}^2$ ) is, the bigger the  $F$  test value  $F_i$  would be. In this situation, the  $F$  test can be passed easily. And it also expresses the attribute effects on the comprehensive evaluation result  $y$  significantly.

Under the condition of the hypothesis of  $H_0$ ,  $F_i$  follows  $F$  distribution; that is to say,  $F_i \sim F(1, n - (m - 2))$ . Let the confidence level  $\alpha$  be equal to 0.05 [45]; the critical value  $F_\alpha$  can be checked from  $F$  statistics. If  $F_i \geq F_\alpha$ , accept hypothesis  $H_0: \Delta R^2 \neq 0$ . It means that  $\Delta R^2$  is not equal to zero significantly, and the attribute  $x_2^*$  should be reserved. Conversely, if  $F_i < F_\alpha$ , reject hypothesis  $H_0: \Delta R^2 \neq 0$ ,

which indicates that  $\Delta R^2$  is equal to zero significantly, and the attribute  $x_2^*$  should be deleted.

*Step 7.* Repeat *Step 3* to *Step 6*, and select other attributes.

For the rest of the  $m-2$  attributes  $x_3^*, x_4^*, \dots, x_m^*$ , we can reduce attributes by repeating *Step 3* to *Step 6*. Until you find the first attribute  $x_i^*$ , the corresponding  $F$  test value satisfies the inequation  $F_i^*(x_i^*) \leq F_{0.05}(x_i^*)$ . At this time, the attribute reduction can be stopped. It suggests that the rest of attributes do not have significant influence on comprehensive evaluation result  $y$ .

*2.4. The Judgment of Reasonability of the Proposed Attribute Reduction Approach.* According to the idea that the multiple determination coefficient  $R^2$  can be used to describe the explanation ability of the independent variable on the dependent variable, this paper uses an information contribution ratio to assess the performance of attribute reduction model. The information contribution ratio can be defined as the ratio of the explanation ability  $R_{\text{Reserved}}^2$  of the reserved attributes to the comprehensive evaluation score  $y$  to the explanation ability  $R_{\text{Mass-election}}^2$  of the mass-election attributes to the comprehensive evaluation score  $y$ .

Let  $\text{In}$  denote an information contribution ratio of the reserved attributes to the mass-election attributes, let  $R_{\text{Reserved}}^2$  denote the multiple determination coefficient of the reserved attributes to the comprehensive evaluation score  $y$ , and let  $R_{\text{Mass-election}}^2$  denote the multiple determination coefficient of the mass-election attributes to the comprehensive evaluation score  $y$ . The information contribution rate  $\text{In}$  of the reserved attributes to the mass-election attributes is given by

$$\text{In} = \frac{R_{\text{Reserved}}^2}{R_{\text{Mass-election}}^2}. \quad (15)$$

Equation (15) is applied to judge the reasonability of the proposed attribute reduction model. The numerator  $R_{\text{Reserved}}^2$  reflects the explanation ability of the reserved attributes to the comprehensive evaluation score  $y$ , and the denominator  $R_{\text{Mass-election}}^2$  illustrates the explanation ability of the mass-election attributes to the comprehensive evaluation score  $y$ . Equation (15) is the ratio of the explanation ability  $R_{\text{Reserved}}^2$  to the explanation ability  $R_{\text{Mass-election}}^2$ . It reveals the information contribution degree of the reserved attributes to the mass-election attributes.

As a decision criterion for judging the rationality of the proposed attribute reduction model, the proposed approach is considered reasonable if the reserved attributes are able to contribute more than 90% of the mass-election attributes by using less than 30% of attributes in the mass-election attribute set.

### 3. Empirical Study

*3.1. Sample Selection and Data Sources.* In consideration of research purpose of verifying the applicability of the proposed attribute reduction model, this subsection implements

empirical study based on the financing ability data of 713 small enterprises. In order to guarantee the representation of empirical results, this paper collected the data from the headquarter and all of the branches in a city commercial bank of China, including Beijing Branch, Tianjin Branch, Shanghai Branch, Chongqing Branch, Shenyang Branch, Dalian Branch, and Dandong Branch. The data is shown in Column 5 to Column 717 in Table 1 [46].

The mass-election attribute set for small enterprise financing ability evaluation contains six criterion layers:  $C_1$ , *enterprise basic situation*;  $C_2$ , *debt paying ability*;  $C_3$ , *enterprise profitability*;  $C_4$ , *operation ability*;  $C_5$ , *development potential*;  $C_6$ , *enterprise external macroconditions*, as shown in Column 2 in Table 1. All of the 69 attributes are listed in Column 3 in Table 1. As known from the fourth Column of Table 1, there are 46 positive attributes, 7 negative attributes, 2 interval attributes, and 14 qualitative attributes.

**3.2. The Attribute Data Standardization.** In this paper, we have two interval attributes: “ $X_4$  the age of enterprise legal person” and “ $X_{67}$  consumer price index (CPI).” The ideal range of “ $X_4$  the age of enterprise legal person” is [31, 45] [25]. It means if the age of the business owner is within the interval [31, 45], the repayment ability and repayment willingness of the small enterprise are strong. The ideal range of “ $X_{67}$  consumer price index (CPI)” is [101, 105] [25]. It indicates that there exists neither deflation nor inflation, when the CPI is within the range [101, 105].

The data standardization for quantitative attributes is as follows: in terms of the attribute type in Column 4 of Table 1, substituting the original data of positive attributes  $v_{ij}$  from Column 5 to 717 of Table 1 into (1), the original data of negative attributes  $v_{ij}$  into (2), and the original data of interval attributes  $v_{ij}$  into (3), the standardized data  $p_{ij}$  of attributes are obtained. The results are shown in Column 718 to 1430 of Table 1.

Subsequently, we will compute the standardized score for the qualitative attributes. Learning from a commercial bank nonfinancial attributes scoring standard [46], the scoring standard of qualitative attributes can be obtained by rational analysis, as shown in Table 2. Then, the standardized scores of qualitative attributes are obtained combined with the attribute type in Column 4 of Table 1, as shown in Column 718 to 1430 of Table 1.

**3.3. Attribute Reduction Utilizing Pearson Correlation Analysis.** In practice, due to the presence of related attribute values but independent attributes of meaning, some of the attributes might be mistakenly deleted. This paper calculates attributes’ Pearson correlation coefficients in the same criterion layer. In order to explain the process of Pearson correlation analysis, we take the 10 attributes of the fourth criterion layer “ $C_4$  Operation ability” as an example.

After substituting the data from Row 45 to 54 and Column 718 to 1430 of Table 1 into (4), the correlation coefficients can be obtained for any two attributes, as shown in Table 3. Known from Table 3, the correlation coefficient 0.998 between “ $X_{53}$  accounts payable turnover speed” and

“ $X_{54}$  cash cycle” is greater than the threshold value 0.8, which means that the two attributes reflect information highly repetitively. Because there are other attributes representing cash flow in the attribute set, such as “ $X_{17}$  the main business income cash ratio” and “ $X_{20}$  all assets cash recovery rate,” we delete the attribute “ $X_{54}$  cash cycle.”

Similarly, we can obtain the attributes’ Pearson correlation coefficients for the rest of five criterion layers. Next, we delete the other 15 attributes:  $X_6$ ,  $X_7$ ,  $X_9$ ,  $X_{11}$ ,  $X_{16}$ ,  $X_{18}$ ,  $X_{21}$ ,  $X_{22}$ ,  $X_{26}$ ,  $X_{31}$ ,  $X_{32}$ ,  $X_{38}$ ,  $X_{40}$ ,  $X_{42}$ , and  $X_{43}$ . The removed 16 attributes are marked with “delete by correlation analysis” in the last column of Table 1.

There are 53 attributes after reducing by Pearson correlation analysis, and the corresponding attributes’ standard data are listed in Column 3 to 55 in Table 4.

**3.4. Attribute Reduction Using F Test Significance Discrimination.** Taking the data of Table 4 into (5) to (7), the entropy weights of 53 attributes can be obtained:

$$w = (0.0178, 0.0102, 0.0030, \dots, 0.0003)_{53 \times 1}^T. \quad (16)$$

Substituting the data of the first row in Table 4 and the entropy weights of 53 attributes into (8), then the comprehensive score  $y_1 = 0.155$  of enterprise 1 can be calculated. Similarly, we can calculate the rest of 712 enterprises’ comprehensive scores  $y_i$ , as shown in the last column of Table 4.

After taking the data from Table 4 into (4), 53 Pearson correlation coefficients between the 53 attributes  $X_i$  and the comprehensive score  $y$  can be obtained. According to the correlation coefficient absolute value  $|r_i|$  in a descending order, the attributes’ ranking results are listed in Table 5. Obviously, the attribute  $X_{61}$  has the maximum correlation coefficient with the comprehensive score  $y$ ; therefore the attribute  $X_{61}$  should be reserved.

In terms of Step 3 in Section 2.3.2, substituting the 52 attributes’ data after removing the attribute  $X_{61}$  and the comprehensive score  $y$  into (9) and (10), then the multiple determination coefficient  $R_{52}^2 = 0.938$  between the comprehensive index  $y$  and the remaining 52 attributes can be got. In a similar way, we can calculate the multiple determination coefficient  $R_{51}^2 = 0.824$  between the comprehensive index  $y$  and the remaining 51 attributes after removing the two attributes  $X_{61}$  and  $X_{63}$ . Then  $\Delta R^2 = R_{52}^2 - R_{51}^2 = 0.938 - 0.824 = 0.114$ . Take  $\Delta R^2 = 0.114$ ,  $R_{m-1}^2 = R_{52}^2 = 0.938$ , and  $n - (m - 2) = 713 - 51 = 662$  into (14);  $F(X_{63}) = 1217.226$  can be obtained.

Let the confidence level  $\alpha$  be equal to 0.05; the critical value  $F_\alpha(1, n - (m - 2)) = F_{0.05}(1, 662) = 3.8556$  can be checked from F statistics. Because  $F(X_{63}) = 1217.226 > F_{0.05}(1, 662) = 3.8556$ , we should accept the hypothesis  $H_0 : \Delta R^2 \neq 0$ . It means that  $\Delta R^2$  is not equal to zero significantly, and the attribute  $X_{63}$  affects small enterprise financing ability evaluation significantly, and therefore it should be reserved.

TABLE 1: Financing ability evaluation attribute data of 713 small enterprises.

(1) No.	(2) Criterion layer	(3) Attribute	(4) Attribute types	Original data $v_{ij}$			Standard data $p_{ij}$			(1431) Reduction results
				(5) Enterprise 1	(717) Enterprise 713	(718) Enterprise 713	(1430) Enterprise 713			
I	$C_1$ enterprise basic situation	$X_1$ the education background of enterprise legal person ...	Qualitative ...	1.000 ...	1.000 ...	1.000 ...	1.000 ...	1.000 ...	Reserve ...	
II		$X_{11}$ the number of contract breaches between enterprises	Qualitative	1.000	1.000	1.000	1.000	1.000	Delete by correlation analysis	
12	$C_2$ debt paying ability	$X_{12}$ asset liability ratio ...	Negative ...	0.501 ...	0.463 ...	0.474 ...	0.513 ...	0.485 ...	Delete by $F$ test significance discrimination ...	
32		$X_{32}$ the ratio of EBITDA and total debt	Positive	0.084	-0.115	0.488	0.485	0.485	Delete by correlation analysis	
33	$C_3$ enterprise profitability	$X_{33}$ net assets income rate ...	Positive ...	0.085 ...	-0.097 ...	0.495 ...	0.440 ...	0.440 ...	Delete by $F$ test significance discrimination ...	
44		$X_{44}$ net cash flow from operating activities	Positive	0.000	-2404133.970	0.503	0.492	0.492	Delete by $F$ test significance discrimination	
45	$C_4$ operation ability	$X_{45}$ Accounts receivable turnover speed ...	Positive ...	4.450 ...	45.060 ...	0.469 ...	0.519 ...	0.519 ...	Delete by $F$ test significance discrimination ...	
54		$X_{54}$ cash cycle	Positive	3.386	23.695	0.502	0.507	0.507	Delete by correlation analysis	
55	$C_5$ development potential	$X_{55}$ business revenue growth rate ...	Positive ...	-0.540 ...	0.030 ...	0.463 ...	0.483 ...	0.483 ...	Delete by $F$ test significance discrimination ...	
63		$X_{63}$ brand product level	Qualitative	0.000	0.750	0.000	0.750	0.750	Reserve	
64	$C_6$ enterprise external macro conditions	$X_{64}$ industry cycle index ...	Positive ...	123.200 ...	135.300 ...	0.578 ...	0.700 ...	0.700 ...	Reserve ...	
69		$X_{69}$ Engel coefficient	Negative	44.000	37.000	0.450	0.755	0.755	Delete by $F$ test significance discrimination	

TABLE 2: The scoring criteria of the qualitative attributes.

(1) No.	(2) Criterion layer	(3) Attribute	(4) Options and its description	(5) Score
1			Bachelor degree or above	1.000
2		$X_1$ , the education background of enterprise legal person	Junior college graduate	0.900
3	$C_1$ enterprise basic situation		Senior high school graduate or education technical secondary school graduate	0.700
4			Junior middle school graduate	0.400
5			Others or data missing	0.000
...	...	...	...	...
63			The main products of the enterprise are the national brand products	1.000
64			The main products of the enterprise are the provincial brand products	0.750
65	$C_5$ development potential	$X_{63}$ brand product level	The main products of the enterprise are the prefecture-level brand products	0.500
67			The main products of the enterprise are the district brand products	0.250
68			Others or data missing	0.000



TABLE 3: The correlation coefficients between attributes for “ $C_4$  operation ability”.

	$X_{45}$	$X_{46}$	$X_{47}$	$X_{48}$	$X_{49}$	$X_{50}$	$X_{51}$	$X_{52}$	$X_{53}$	$X_{54}$
$X_{45}$	1.000									
$X_{46}$	0.425	1.000								
$X_{47}$	0.371	0.297	1.000							
$X_{48}$	0.349	0.297	0.783	1.000						
$X_{49}$	0.043	0.009	0.064	-0.010	1.000					
$X_{50}$	0.249	0.088	0.513	0.486	0.070	1.000				
$X_{51}$	0.016	0.046	0.043	-0.054	-0.045	-0.139	1.000			
$X_{52}$	0.004	0.005	0.063	0.003	0.006	-0.024	0.030	1.000		
$X_{53}$	-0.146	-0.185	-0.023	-0.058	-0.032	-0.030	-0.119	0.027	1.000	
$X_{54}$	-0.098	-0.179	0.002	-0.037	-0.030	-0.020	-0.121	0.029	<b>0.998</b>	1.000

TABLE 4: 53 attributes and the comprehensive score  $y_i$  of 713 small enterprises.

(1) No.	(2) Enterprise	(3) $X_1$	...	(46) $X_{60}$	(47) $X_{61}$	(48) $X_{62}$	(49) $X_{63}$	...	(55) $X_{69}$	(56) Comprehensive score $y_i$
1	Enterprise 1	1.000	...	0.000	<b>0.000</b>	1.000	<b>0.000</b>	...	0.450	0.155
2	Enterprise 2	0.000	...	0.000	<b>0.000</b>	1.000	<b>0.000</b>	...	0.790	0.177
3	Enterprise 3	1.000	...	0.250	<b>0.000</b>	1.000	<b>0.000</b>	...	0.755	0.176
4	Enterprise 4	0.000	...	0.000	<b>0.000</b>	1.000	<b>0.000</b>	...	0.790	0.177
5	Enterprise 5	1.000	...	0.250	<b>0.000</b>	1.000	<b>0.000</b>	...	0.755	0.176
...	...	...	...	...	...	...	...	...	...	...
711	Enterprise 711	1.000	...	0.250	<b>0.500</b>	0.500	<b>0.250</b>	...	0.751	0.314
712	Enterprise 712	1.000	...	0.250	<b>0.000</b>	0.500	<b>0.000</b>	...	0.751	0.263
713	Enterprise 713	1.000	...	0.750	<b>0.500</b>	0.500	<b>0.750</b>	...	0.755	0.544

Similarly, after repeating this process as mentioned above, we can select other 17 attributes, as shown in Row 3 to Row 19 of Table 5. Until the 20th attribute  $X_{12}$ , the  $F$  test value  $F(X_{12})$  equals 2.3591. And the corresponding  $F_{0.05}(1, 681)$  equals 3.8552. Obviously,  $F(X_{12}) = 2.3591 < F_{0.05}(1, 681) = 3.8552$ ; thus we should accept the hypothesis  $H_1 : \Delta R^2 = 0$ . It indicates that the attribute  $X_{12}$  does not have significant influence on the comprehensive evaluation result  $y$ , and it should be deleted. At this time, the attribute reduction can be stopped according to Step 7 in Section 2.3.2.

The selected 19 attributes are marked with “*reserve*” in the last column of Table 5. And the deleted 34 attributes are marked with “*delete by F test significance discrimination*” in the last column of Table 5. The selected 19 key influencing factors of small enterprise financing ability are shown in Table 6. And the detailed attribute reduction process of financing ability evaluation for 713 small enterprises is shown in Table 7.

**3.5. The Reasonability Judgment for the Proposed Attribute Reduction Model.** In Table 4, taking the data of 19 reserved attributes and the comprehensive evaluation score  $y$  into (9) and (10), the multiple determination coefficient  $R_{\text{Reserved}}^2 =$

0.947 can be obtained. In Table 1, taking the data of 69 original attributes and the comprehensive evaluation score  $y$  into (9) and (10), the multiple determination coefficient  $R_{\text{Mass-election}}^2 = 1.000$  can be obtained. Thus  $\text{In} = R_{\text{Reserved}}^2 / R_{\text{Mass-election}}^2 = 94.7\%$ . It illustrates that the selected attributes reflect 94.7% original information with 27.54% attributes (27.54% = 19/69) by using the proposed attribute reduction model. And the experimental results show that the proposed reduction model is efficient and effective.

**3.6. Some Notes about the Proposed Model.** In Section 2.3.1, this paper takes entropy weight method as an example for the purpose of illustrating the feasibility and rationality of the proposed attribute reduction idea. As a matter of fact, the weight methods can be substituted in terms of the needs of decision makers. They can select other weight methods, such as AHP, G1, G2, and interval numbers weight approaches [47].

In Section 2.3.2, the paper takes linear regression model as an example so as to explain the feasibility of the proposed model. In reality, decision makers can select other nonlinear regression models [48].

TABLE 5: Attribute reduction process based on  $F$  test significance discrimination.

(1) No.	(2) $X_i$	(3) Correlation coefficient between attribute $X_i$ and $y$	(4) Relevance ranking between attribute $X_i$ and $y$	(5) $\Delta R^2$	(6) $F(X_i)$	(7) $F_{0.05}(1, n - (m - 2))$	(8) Reduction results
1	$X_{61}$	0.782	1	0.114	1217.2	3.8556	Reserve
2	$X_{63}$	0.721	2	0.351	1322.2	3.8555	Reserve
3	$X_{60}$	0.717	3	0.033	41.579	3.8555	Reserve
...	...	...	...	...	...	...	...
19	$X_3$	0.136	19	0.01	7.9439	3.8552	Reserve
20	$X_{12}$	0.126	20	0.003	2.3591	3.8552	Delete by $F$ test significance discrimination
...	...	...	...	...	...	...	...
53	$X_{14}$	0.000	53	—	—	—	Delete by $F$ test significance discrimination

TABLE 6: The key influencing factors of small enterprise financing ability.

(1) No.	(2) Criterion layer	(3) Influencing factors
1	C <sub>1</sub> enterprise basic situation	X <sub>1</sub> the education background of enterprise legal person
2		X <sub>2</sub> local residence permission of enterprise legal person
3		X <sub>3</sub> the gender of enterprise legal person
4		X <sub>5</sub> years for the job of enterprise legal person
5		X <sub>8</sub> enterprise tax records
6		X <sub>10</sub> enterprise law-abiding business conditions
7	C <sub>2</sub> debt paying ability	X <sub>15</sub> quick ratio
8		X <sub>17</sub> the main business income cash ratio
9		X <sub>19</sub> full capitalization rate
10		X <sub>30</sub> noncurrent liabilities business activities net cash flow ratio
11	C <sub>3</sub> enterprise profitability	X <sub>36</sub> total assets return rate
12		X <sub>39</sub> gross profit rate
13		X <sub>41</sub> EBITDA
14	C <sub>4</sub> operation ability	X <sub>52</sub> rate of return on investment
15	C <sub>5</sub> development potential	X <sub>60</sub> new product identification level
16		X <sub>61</sub> patent status
17		X <sub>62</sub> product sales scope
18		X <sub>63</sub> brand product level
19	C <sub>6</sub> enterprise external macroconditions	X <sub>64</sub> industry cycle index

#### 4. Conclusions and Future Work

In order to reduce large dimensionality in complex data sets, we create an attribute reduction approach based on Pearson correlation analysis and  $F$  test significance discrimination. First of all, we delete redundancy attributes using Pearson correlation coefficient, avoiding information chaos of the original attribute data sets. Secondly, developing attribute reduction methodology utilizing  $F$  test significance discrimination can find the key attributes that have the greatest influence on the evaluation results. Thirdly, the paper also defines an information contribution ratio to assess the performance of attribute reduction model from a statistical viewpoint.

The proposed attribute reduction model has been verified utilizing the financing ability evaluation data of 713 small enterprises of a city commercial bank in China. The empirical evidence shows the accuracy and applicability of the proposed model. Moreover, we also establish an evaluation indicator system for small enterprise financing ability. It will help the downstream organizations of supply chain to choose more qualified partners and alleviate the difficulties faced by enterprises when applying loan. Furthermore, applications of the proposed model to real world data are expected in future.

It is well known that the problems of attribute reduction are ubiquitous in data mining activities. The empirical study

in this paper is only an example in order to verify the accuracy of the proposed model. A topic of future research can be the application of the proposed approach to data sets in other attribute reduction areas. Researchers can easily conduct attribute reduction through cases and empirical studies.

#### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of the paper.

#### Acknowledgments

The study was supported by the National Natural Science Foundation of China (nos. 71503199, 71471027, 71403215, 71373207, 71703122), the Key Project of National Natural Science Foundation of China (no. 71731003), the Project Funded by the China Postdoctoral Science Foundation (nos. 2016T90957, 2016M600209, 2015M572608), the Science and Technology Plan Project of Yangling Demonstration Zone (no. 2016RKX-04), and the Credit Rating and Loan Pricing Project for Small Enterprise of Bank of Dalian (no. 2012-01). Special thanks go to the Youth Talent Cultivation Program of Northwest A&F University (no. Z109021717). The first author would like to thank Professor Chunguang Bai from Dongbei

TABLE 7: The attribute reduction process of financing ability evaluation for 713 small enterprises.

(1) No.	(2) Criterion layer	(3) Attribute	(4) Attribute types	Original data				Standard data				(1431) Reduction results
				(5) Enterprise 1	...	Enterprise 713	(717) Enterprise 1	(718) Enterprise 1	...	(1430) Enterprise 713		
1	C <sub>1</sub> enterprise basic situation	X <sub>1</sub> the education background of enterprise legal person	Qualitative	1.000	...	1.000	1.000	1.000	1.000	...	1.000	Reserve
2		X <sub>2</sub> local residence permission of enterprise legal person	Qualitative	1.000	...	1.000	1.000	1.000	1.000	...	1.000	Reserve
3		X <sub>3</sub> the gender of enterprise legal person	Qualitative	1.000	...	1.000	1.000	1.000	1.000	...	1.000	Reserve
4		X <sub>4</sub> the age of enterprise legal person	Interval	36.000	...	40.000	1.000	1.000	1.000	...	1.000	Delete by <i>F</i> test significance discrimination
5		X <sub>5</sub> years for the job of enterprise legal person	Qualitative	0.400	...	1.000	0.400	0.400	1.000	...	1.000	Reserve
6		X <sub>6</sub> enterprise registered capital category	Qualitative	1.000	...	1.000	1.000	1.000	1.000	...	1.000	Delete by correlation analysis
7		X <sub>7</sub> enterprise credit, nearly three years	Qualitative	1.000	...	1.000	1.000	1.000	1.000	...	1.000	Delete by correlation analysis
8		X <sub>8</sub> enterprise tax records	Qualitative	1.000	...	1.000	1.000	1.000	1.000	...	1.000	Reserve
9		X <sub>9</sub> enterprise legal disputes	Qualitative	1.000	...	1.000	1.000	1.000	1.000	...	1.000	Delete by correlation analysis
10		X <sub>10</sub> enterprise law-abiding business conditions	Qualitative	0.500	...	1.000	0.500	0.500	1.000	...	1.000	Reserve
11	X <sub>11</sub> the number of contract breaches between enterprises	Qualitative	1.000	...	1.000	1.000	1.000	1.000	...	1.000	Delete by correlation analysis	
12	C <sub>2</sub> debt paying ability	X <sub>12</sub> asset liability ratio	Negative	0.501	...	0.463	0.474	0.474	...	0.513	Delete by <i>F</i> test significance discrimination	
13		X <sub>13</sub> current liabilities business activities net cash flow ratio	Positive	0.000	...	-0.024	0.492	0.492	...	0.492	Delete by <i>F</i> test significance discrimination	
14		X <sub>14</sub> interest safeguard multiple	Positive	0.000	...	0.000	0.440	0.440	...	0.440	Delete by <i>F</i> test significance discrimination	
15		X <sub>15</sub> quick ratio	Positive	0.200	...	0.320	0.457	0.457	...	0.459	Reserve	
16		X <sub>16</sub> current ratio	Positive	0.830	...	0.600	0.471	0.471	...	0.469	Delete by correlation analysis	
17		X <sub>17</sub> the main business income cash ratio	Positive	1.240	...	1.000	0.694	0.694	...	0.625	Reserve	
18		X <sub>18</sub> profit current liabilities ratio before interest and tax	Positive	0.084	...	-0.115	0.489	0.489	...	0.486	Delete by correlation analysis	
19		X <sub>19</sub> full capitalization rate	Negative	0.000	...	0.000	1.000	1.000	...	1.000	Reserve	
20		X <sub>20</sub> all assets cash recovery rate	Positive	0.000	...	-0.011	0.475	0.475	...	0.466	Delete by <i>F</i> test significance discrimination	
21		X <sub>21</sub> shareholders' equity ratio	Positive	0.499	...	0.537	0.474	0.474	...	0.513	Delete by correlation analysis	
22	X <sub>22</sub> superquick ratio	Positive	0.097	...	0.288	0.476	0.476	...	0.479	Delete by correlation analysis		
23	X <sub>23</sub> net profit cash content	Positive	0.000	...	0.210	0.496	0.496	...	0.496	Delete by <i>F</i> test significance discrimination		
24	X <sub>24</sub> the ratio of net assets and year-end loan balance	Positive	1.970	...	2.350	0.493	0.493	...	0.494	Delete by <i>F</i> test significance discrimination		

TABLE 7: Continued.

(1) No.	(2) Criterion layer	(3) Attribute	(4) Attribute types	Original data			Standard data			(1431) Reduction results
				(5) Enterprise 1	...	(717) Enterprise 713	(718) Enterprise 1	...	(1430) Enterprise 713	
25		$X_{25}$ capital fixed ratio	Negative	1.170	...	1.345	0.485	...	0.479	Delete by $F$ test significance discrimination
26		$X_{26}$ Cash ratio	Positive	0.010	...	0.240	0.468	...	0.474	Delete by correlation analysis
27		$X_{27}$ long-term assets appropriate rate	Positive	0.850	...	0.740	0.467	...	0.467	Delete by $F$ test significance discrimination
28	$C_2$ debt paying ability	$X_{28}$ total outstanding loans accounted for the proportion of net assets	Negative	0.000	...	0.000	0.530	...	0.497	Delete by $F$ test significance discrimination
29		$X_{29}$ total outstanding loans accounted for the proportion of general assets	Negative	0.000	...	0.000	1.000	...	0.950	Delete by $F$ test significance discrimination
30		$X_{30}$ noncurrent liabilities business activities net cash flow ratio	Positive	-155.323	...	-155.323	0.000	...	0.000	Reserve
31		$X_{31}$ total liabilities business activities net cash flow ratio	Positive	0.000	...	-0.024	0.492	...	0.492	Delete by correlation analysis
32		$X_{32}$ the ratio of EBITDA and total debt	Positive	0.084	...	-0.115	0.488	...	0.485	Delete by correlation analysis
33		$X_{33}$ net assets income rate	Positive	0.085	...	-0.097	0.495	...	0.440	Delete by $F$ test significance discrimination
34		$X_{34}$ sales net rate	Positive	0.000	...	-0.012	0.506	...	0.505	Delete by $F$ test significance discrimination
35		$X_{35}$ sales net profit rate	Positive	0.098	...	-0.059	0.716	...	0.693	Delete by $F$ test significance discrimination
36		$X_{36}$ total assets return rate	Positive	0.070	...	-0.030	0.497	...	0.450	Reserve
37	$C_3$ enterprise profitability	$X_{37}$ operating profit rate	Positive	0.100	...	-0.060	0.563	...	0.539	Delete by $F$ test significance discrimination
38		$X_{38}$ cost net interest rate	Positive	0.109	...	-0.056	0.497	...	0.484	Delete by correlation analysis
39		$X_{39}$ gross profit rate	Positive	0.320	...	0.010	0.822	...	0.741	Reserve
40		$X_{40}$ cost-profit ratio	Positive	0.109	...	-0.057	0.495	...	0.482	Delete by correlation analysis
41		$X_{41}$ EBITDA	Positive	1016249.670	...	-11686752.290	0.658	...	0.599	Reserve
42		$X_{42}$ EBITDA profit rate	Positive	0.098	...	-0.060	0.509	...	0.492	Delete by correlation analysis
43		$X_{43}$ net profit	Positive	1016249.670	...	-11451533.390	0.661	...	0.604	Delete by correlation analysis
44		$X_{44}$ net cash flow from operating activities	Positive	0.000	...	-2404133.970	0.503	...	0.492	Delete by $F$ test significance discrimination



TABLE 7: Continued.

(1) No.	(2) Criterion layer	(3) Attribute	(4) Attribute types	Original data			Standard data			(1431) Reduction results
				(5) Enterprise 1	...	(717) Enterprise 713	(718) Enterprise 1	...	(1430) Enterprise 713	
45	$X_{45}$	accounts receivable turnover speed	Positive	4.450	...	45.060	0.469	...	0.519	Delete by $F$ test significance discrimination
46	$X_{46}$	inventory turnover speed	Positive	1.150	...	6.880	0.478	...	0.482	Delete by $F$ test significance discrimination
47	$X_{47}$	total assets turnover speed	Positive	0.480	...	0.890	0.350	...	0.400	Delete by $F$ test significance discrimination
48	$X_{48}$	current assets turnover speed	Positive	1.038	...	3.207	0.428	...	0.648	Delete by $F$ test significance discrimination
49	$C_4$	operation ability	Positive	0.797	...	1.308	0.484	...	0.484	Delete by $F$ test significance discrimination
50		shareholders' equity turnover speed	Positive	0.866	...	1.656	0.443	...	0.455	Delete by $F$ test significance discrimination
51	$X_{51}$	working capital allocation ratio	Positive	-0.204	...	-0.668	0.977	...	0.969	Delete by $F$ test significance discrimination
52	$X_{52}$	rate of return on investment	Positive	0.000	...	0.000	0.490	...	0.490	Reserve
53	$X_{53}$	accounts payable turnover speed	Negative	2.214	...	28.245	0.791	...	0.780	Delete by $F$ test significance discrimination
54	$X_{54}$	cash cycle	Positive	3.386	...	23.695	0.502	...	0.507	Delete by correlation analysis
55	$X_{55}$	business revenue growth rate	Positive	-0.540	...	0.030	0.463	...	0.483	Delete by $F$ test significance discrimination
56	$X_{56}$	profit growth rate	Positive	-0.620	...	0.160	0.495	...	0.496	Delete by $F$ test significance discrimination
57	$X_{57}$	total assets growth rate	Positive	0.024	...	0.020	0.471	...	0.471	Delete by $F$ test significance discrimination
58	$C_5$	development potential	Positive	3.350	...	-1.780	0.494	...	0.494	Delete by $F$ test significance discrimination
59		retained earnings growth rate	Positive	0.008	...	5.781	0.494	...	0.498	Delete by $F$ test significance discrimination
60	$X_{60}$	new product identification level	Qualitative	0.000	...	0.750	0.000	...	0.750	Reserve
61	$X_{61}$	patent status	Qualitative	0.000	...	0.500	0.000	...	0.500	Reserve
62	$X_{62}$	product sales scope	Qualitative	1.000	...	0.500	1.000	...	0.500	Reserve
63	$X_{63}$	brand product level	Qualitative	0.000	...	0.750	0.000	...	0.750	Reserve

TABLE 7: Continued.

(1) No.	(2) Criterion layer	(3) Attribute	(4) Attribute types	Original data			Standard data			(1431) Reduction results
				(5) Enterprise 1	(717) Enterprise 713	(718) Enterprise 1	(1430) Enterprise 713	(1431) Enterprise 713		
64		$X_{64}$ industry cycle index	Positive	123.200	135.300	0.578	...	0.700	Reserve	
65		$X_{65}$ urban residents per capita savings yearend balance	Positive	31965.990	44133.430	0.343	...	0.491	Delete by $F$ test significance discrimination	
66	$C_6$ enterprise external macroconditions	$X_{66}$ the GDP growth rate	Positive	21.900	15.000	0.697	...	0.346	Delete by $F$ test significance discrimination	
67		$X_{67}$ consumer price index (CPI)	Interval	105.100	103.000	0.994	...	1.000	Delete by $F$ test significance discrimination	
68		$X_{68}$ urban residents per capita disposable income	Positive	17123.380	19920.000	0.424	...	0.508	Delete by $F$ test significance discrimination	
69		$X_{69}$ Engel coefficient	Negative	44.000	37.000	0.450	...	0.755	Delete by $F$ test significance discrimination	

University of Finance and Economics for her valuable comments and suggestions.

## References

- [1] J. Manyika, M. Chui et al., *Big Data: The Next Frontier for Innovation, Competition, And Productivity*, vol. 4, McKinsey Global Institute, 2011.
- [2] C. Bai and J. Sarkis, "Determining and applying sustainable supplier key performance indicators," *Supply Chain Management Review*, vol. 19, no. 3, pp. 275–291, 2014.
- [3] B. Li, T. W. S. Chow, and P. Tang, "Analyzing rough set based attribute reductions by extension rule," *Neurocomputing*, vol. 123, pp. 185–196, 2014.
- [4] H. Li, G. Chen, T. Huang, and Z. Dong, "High-performance consensus control in networked systems with limited bandwidth communication and time-varying directed topologies," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2016.
- [5] G. Chen, Z. Y. Dong, D. J. Hill, G. H. Zhang, and K. Q. Hua, "Attack structural vulnerability of power grids: a hybrid approach based on complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 3, pp. 595–603, 2010.
- [6] S.-H. Teng, M. Lu, A.-F. Yang, J. Zhang, Y. Nian, and M. He, "Efficient attribute reduction from the viewpoint of discernibility," *Information Sciences*, vol. 326, pp. 297–314, 2016.
- [7] Z. Dong, M. Sun, and Y. Yang, "Fast algorithms of attribute reduction for covering decision systems with minimal elements in discernibility matrix," *International Journal of Machine Learning and Cybernetics*, vol. 7, no. 2, pp. 297–310, 2016.
- [8] Z. a. Pawlak, "Rough sets," *International Journal of Computer & Information Science*, vol. 11, no. 5, pp. 341–356, 1982.
- [9] M. T. Rezvan, A. Z. Hamadani, and S. R. Hejazi, "An exact feature selection algorithm based on rough set theory," *Complexity*, vol. 20, no. 5, pp. 50–62, 2015.
- [10] B. Shi, J. Zhao, and J. Wang, "A credit rating attribute reduction approach based on pearson correlation analysis and fuzzy-rough sets," *ICIC Express Letters*, vol. 10, no. 2, pp. 519–525, 2016.
- [11] C. Degang, W. Changzhong, and H. Qinghua, "A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets," *Information Sciences*, vol. 177, no. 17, pp. 3500–3518, 2007.
- [12] Q. Hu, Z. Xie, and D. Yu, "Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation," *Pattern Recognition*, vol. 40, no. 12, pp. 3509–3521, 2007.
- [13] X. Yang, J. Yang, C. Wu, and D. Yu, "Dominance-based rough set approach and knowledge reductions in incomplete ordered information system," *Information Sciences*, vol. 178, no. 4, pp. 1219–1234, 2008.
- [14] S. Greco, B. Matarazzo, R. Slowinski, and J. Stefanowski, "Variable consistency model of dominance-based rough sets approach," in *Rough Sets and Current Trends in Computing*, W. Ziarko and Y. Yao, Eds., vol. 2005 of *Lecture Notes in Computer Science*, pp. 170–181, Springer, Berlin, Germany, 2001.
- [15] M. Inuiguchi, Y. Yoshioka, and Y. Kusunoki, "Variable-precision dominance-based rough set approach and attribute reduction," *International Journal of Approximate Reasoning*, vol. 50, no. 8, pp. 1199–1214, 2009.
- [16] E. C. C. Tsang, C. Degang, and D. S. Yeung, "Approximations and reducts with covering generalized rough sets," *Computers & Mathematics with Applications*, vol. 56, no. 1, pp. 279–289, 2008.
- [17] C. Wang, M. Shao, B. Sun, and Q. Hu, "An improved attribute reduction scheme with covering based rough sets," *Applied Soft Computing*, vol. 26, pp. 235–243, 2015.
- [18] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, 2008.
- [19] A. M. Radzikowska and E. E. Kerre, "A comparative study of fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 126, no. 2, pp. 137–155, 2002.
- [20] C. Bai, J. Sarkis, X. Wei, and L. Koh, "Evaluating ecological sustainable performance measures for supply chain management," *Supply Chain Management Review*, vol. 17, no. 1, pp. 78–92, 2012.
- [21] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *Parallel Problem Solving from Nature—PPSN VIII*, vol. 3242 of *Lecture Notes in Computer Science*, pp. 832–842, Springer, Berlin, Germany, 2004.
- [22] K. Polat and V. Krmac, "Determining of gas type in counter flow vortex tube using pairwise fisher score attribute reduction method," *International Journal of Refrigeration*, vol. 34, no. 6, pp. 1372–1386, 2011.
- [23] Y.-H. Ju and S.-Y. Sohn, "Updating a credit-scoring model based on new attributes without realization of actual data," *European Journal of Operational Research*, vol. 234, no. 1, pp. 119–126, 2014.
- [24] R. J. Elliott, T. K. Siu, and E. S. Fung, "A Double HMM approach to Altman Z-scores and credit ratings," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1553–1560, 2014.
- [25] B. F. Shi, J. Wang, J. Y. Qi, and Y. Q. Cheng, "A novel imbalanced data classification approach based on logistic regression and Fisher discriminant," *Mathematical Problems in Engineering*, vol. 2015, Article ID 945359, 12 pages, 2015.
- [26] L. Li and J. Zhang, "Attribute reduction in fuzzy concept lattices based on the T implication," *Knowledge-Based Systems*, vol. 23, no. 6, pp. 497–503, 2010.
- [27] T.-J. Li, M.-Z. Li, and Y. Gao, "Attribute reduction of concept lattice based on irreducible elements," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 11, no. 6, Article ID 1350046, 2013.
- [28] S. M. Dias and N. J. Vieira, "Concept lattices reduction: Definition, analysis and classification," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7084–7097, 2015.
- [29] L. Wei, H.-R. Li, and W.-X. Zhang, "Knowledge reduction based on the equivalence relations defined on attribute set and its power set," *Information Sciences*, vol. 177, no. 15, pp. 3178–3185, 2007.
- [30] F. Min, H. He, Y. Qian, and W. Zhu, "Test-cost-sensitive attribute reduction," *Information Sciences*, vol. 181, no. 22, pp. 4928–4942, 2011.
- [31] G.-T. Chi, T.-T. Cao, and K. Zhang, "The establishment of human all-around development evaluation indicators system based on correlation-principle component analysis," *Xitong Gongcheng Lilun yu Shijian/System Engineering Theory and Practice*, vol. 32, no. 1, pp. 111–119, 2012.
- [32] Z. Xu, H. Zhao, F. Min, and W. Zhu, "Ant colony optimization with three stages for independent test cost attribute reduction," *Mathematical Problems in Engineering*, vol. 2013, Article ID 510167, 2013.
- [33] B.-F. Shi and G.-T. Chi, "Green industry evaluation indicators screening model based on the maximum information content and its application," *Xitong Gongcheng Lilun yu Shijian/System Engineering Theory and Practice*, vol. 34, no. 7, pp. 1799–1810, 2014.

- [34] L. Li, J. Mi, and B. Xie, "Attribute reduction based on maximal rules in decision formal context," *International Journal of Computational Intelligence Systems*, vol. 7, no. 6, pp. 1044–1053, 2014.
- [35] B. Shi, H. Yang, J. Wang, and J. Zhao, "City green economy evaluation: empirical evidence from 15 sub-provincial cities in China," *Sustainability*, vol. 8, no. 6, article 551, 2016.
- [36] K. Pearson, "Note on regression and inheritance in the case of two parents," *Proceedings of The Royal Society of London*, vol. 58, pp. 240–242, 1895.
- [37] Pearson Correlation Coefficients, "Pearson correlation coefficient," [http://en.wikipedia.org/wiki/Pearson\\_product-moment\\_correlation\\_coefficient](http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient).
- [38] A. Buda and A. Jarynowski, "Life time of correlations and its applications," *Wydawnictwo Niezalezne*, p. 21, 2010.
- [39] J. Cohen, *Statistical power analysis for the behavioral sciences*, Routledge Academic, 2013.
- [40] X. Meng, R. Rosenthal, and D. B. Rubin, "Comparing correlated correlation coefficients," *Psychological Bulletin*, vol. 111, no. 1, pp. 172–175, 1992.
- [41] X.-M. Zhang, G.-L. Diao, Y.-P. Zhao, W.-M. Wang, and P.-Y. Shu, "Study on mantle shear wave velocity structures in North China," *Chinese Journal of Geophysics (Acta Geophysica Sinica)*, vol. 49, no. 6, pp. 1709–1719, 2006.
- [42] Y. Liu, L. Zou, Y. Sun, and X. Yang, "Evaluation Model of Aluminum Alloy Welded Joint Low-Cycle Fatigue Data Based on Information Entropy," *Entropy*, vol. 19, no. 1, p. 37, 2017.
- [43] M. Lonnie, " $R^2$  measures based on Wald and likelihood ratio joint significance tests," *The American Statistician*, vol. 44, no. 3, pp. 250–253, 1990.
- [44] L. H. Herbach, "Properties of model II - type analysis of variance tests, A: Optimum nature of the  $F$ -test for model II in the balanced case," *Annals of Mathematical Statistics*, vol. 30, pp. 939–959, 1959.
- [45] E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG- and MEG-data," *Journal of Neuroscience Methods*, vol. 164, no. 1, pp. 177–190, 2007.
- [46] Bank of Dalian (DB), *Small Enterprise Credit Rating System for Bank of Dalian*, Bank of Dalian, 2016.
- [47] B. Meng and G. Chi, "Evaluation index system of green industry based on maximum information content," *The Singapore Economic Review*, pp. 1–20, 2016.
- [48] W. Dillon and M. Goldstein, *Multivariate analysis: Methods and applications*, Wiley Press, New York, NY, USA, 1984.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

