Research Article

A New Knowledge Characteristics Weighting Method Based on Rough Set and Knowledge Granulation

Zhenquan Shi\textsuperscript{1,2} and Shiping Chen\textsuperscript{1} \\

\textsuperscript{1}Business School, University of Shanghai for Science and Technology, Shanghai 200093, China \\
\textsuperscript{2}Nantong University, Nantong, Jiangsu 226017, China

Correspondence should be addressed to Shiping Chen; 56254268@qq.com

Received 12 November 2017; Revised 14 April 2018; Accepted 26 April 2018; Published 31 May 2018

1. Introduction

In data mining, in order to effectively classify the knowledge, we need to make proper assessment on the knowledge characteristics sets. Therefore, it is very important to compute the weights of characteristics sets. Weights reflect the role of characteristics in the classification process and directly affect the validity and accuracy of the classifier. The common weighting methods include experts scoring method, fuzzy statistics method [1–3], Analytic Hierarchy Process (AHP) method [4–6], and Principal Component Analysis (PCA) method [7, 8]. In these methods, the a priori knowledge must be used.

Rough set theory was firstly proposed by Pawlak in 1982 [9]. It has become an extremely useful tool to handle the imprecision and uncertainty knowledge [9, 10]. Rough set theory can be used to analyze and process the fuzzy or uncertain data without the a priori knowledge [11–17]. Now, the rough set theory has been widely used in pattern recognition [18–20], data mining [21–23], machine learning [24–29], and other fields [30–36].

In recent years, the rough set method has been studied to calculate the characteristics weight. For instance, based on the concepts of characteristics importance, Wang et al. proposed a method to determine the characteristics weights. However, this method did not consider the influence of decision characteristics on conditional characteristics [37]. Cao and Liang combined the characteristics importance of the rough set and the experts’ a priori knowledge to determine the characteristics weight [38]. This method achieved the unity of the subjective a priori knowledge with the objective situations, but it ignored the internal difference in the equivalent partitions. Therefore, some nonredundant characteristics would be handled by redundant characteristics. Bao et al. proposed a method ascertaining characteristics weight based on rough set and conditional information entropy. It avoids some nonredundant characteristics to be handled by redundant characteristics. But in this method the characteristics importance obtained by redundant characteristics was higher than that got by nonredundant characteristics [39]. Zhu and Chen constructed the priority queue of characteristics importance to improve Bao’s research. They presented a weighting
method based on the conditional information entropy and rough set, but that method also involved additional costs [40].

In this paper, a new knowledge characteristics weighting method based on the rough set and knowledge granulation theory is proposed. The accuracy of equivalent partitions in knowledge characteristics is studied and the difference in equivalence classes is analyzed. Experimental results on several UCI data sets confirm our theoretical results. By comparing the numerical results with those of the AHP method, the PCA method, and two rough set based methods, we can draw the conclusion that our new method can effectively avoid taking nonredundant characteristics as redundant characteristics and can improve classification accuracy.

The rest of the paper is structured as follows. Some basic concepts about rough set are briefly introduced in Section 2. In Section 3, a new knowledge characteristics weighting method is proposed and studied. Some experimental results are given in Section 4 to show the effectiveness of the proposed weighting method. Finally, we end this paper with some conclusions in Section 5.

2. Basic Concepts

2.1. Rough Set. Rough set theory takes knowledge as a partition of the objects domain. The equivalence relations and equivalence classes produced by the equivalence relations are valid information or knowledge about the objects domain. Let $U$ denote the universe of objects, which is a nonempty set. $\mathcal{R} \subseteq U \times U$ is an equivalence relation, called the knowledge on the universe $U$. The equivalence relation $\mathcal{R}$ divides $U$ into the disjoint subsets; it is denoted as $U/\mathcal{R}$, called the equivalence classes. For the subset $X$ of the universe $U$, there are the equivalence classes $X$. In general, there are two approximation sets—the lower approximation (set) $\mathcal{R}(X) = \{x | [x]_\mathcal{R} \subseteq X\}$ and the upper approximation (set) $\mathcal{R}(X) = \{x | [x]_\mathcal{R} \cap X \neq \phi\}$. The lower approximation (set) of the set $X$ is also defined as the positive region of $\mathcal{R}(X) = \mathcal{R}(X)$. The set $\mathcal{R}(X) = \mathcal{R}(X) - \mathcal{R}(X)$ will be referred to as the $\mathcal{R}$-boundary region of $X$. Obviously, when the border area is larger, the set $X$ divided by $\mathcal{R}$ is rougher. Therefore, the roughness of rough set $X$ about the equivalence relation $\mathcal{R}$ can be achieved; it is denoted by

$$D_R(\mathcal{R}(X), \mathcal{R}(X)) = \frac{|\text{BND}_\mathcal{R}(X)|}{|\mathcal{R}(X)|}. \quad (1)$$

The accuracy of rough set $X$ about the equivalence relation $\mathcal{R}$ is defined as

$$\rho_R(X) = 1 - D_R(\mathcal{R}(X), \mathcal{R}(X)) = \frac{|\mathcal{R}(X)|}{|\mathcal{R}(X)|}, \quad (2)$$

where $|\cdot|$ represents the number of the elements in the collection, $0 \leq \rho_R(X) \leq 1$. When $\rho_R(X) = 1$, $X$ is defined as the accuracy set about the equivalence relation $\mathcal{R}$.

2.2. Knowledge Granularity. By the rough set theory, people learn that knowledge is related to the equivalence classes, which shows that knowledge is granular. That is why some scholars also identify the structure of knowledge granularity by the equivalence classes and calculate the size of the knowledge granularity [39].

Suppose that $K = (U, \mathcal{R})$ is a knowledge base, and $\mathcal{R}$ is an equivalence relation, also known as knowledge. Knowledge granularity is defined as

$$\text{GD}(\mathcal{R}) = \frac{|\mathcal{R}|}{|U|^2}. \quad (3)$$

If the granularity of $\mathcal{R}$ reaches its minimum, then $\text{GD}(\mathcal{R}) = |U|/|U|^2 = 1/|U|$. If $\mathcal{R}$ reaches the universe $U$, i.e., the granularity reaches its maximum, then $\text{GD}(\mathcal{R}) = |U|^2/|U|^2 = 1$. If $(u, v) \in \mathcal{R}$, it indicates that the objects $u$ and $v$ belong to the same equivalence class with the equivalence relation $\mathcal{R}$; they are indiscernible. Obviously, the smaller $\text{GD}(\mathcal{R})$ is, the stronger the discernibility of $\mathcal{R}$ becomes.

Assume that $R$ is an equivalence relation, $K = (U, R)$ is a knowledge base, and $U/R = \{X_1, X_2, \ldots, X_n\}$ is the equivalence class. According to (3), the knowledge granularity can be expressed as

$$\text{GD}(R) = \sum_{i=1}^{n} \frac{|X_i|^2}{|U|^2}. \quad (4)$$

And the discernibility of $R$ is defined as

$$\text{Dis}(R) = 1 - \text{GD}(R). \quad (5)$$

According to (4), there is $\text{Dis}(R) = 1 - \sum_{i=1}^{n} (|X_i|^2/|U|^2)$. Therefore, we have $0 \leq \text{Dis}(R) \leq 1 - 1/|U|$.

3. Knowledge Characteristics Weighting Based on Rough Set and Knowledge Granulation

Cao and Liang calculated the characteristics weights by the cardinality of the positive region set over the cardinality of the discourse set, but the results may be inaccurate [38]. For example, on the field $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, let $X = \{1, 2, 3, 4, 5, 8\}$, and let $R_1$ and $R_2$ be defined as the equivalence relation on $U$. Then the following equivalence classes can be obtained:

$$\frac{U}{R_1} = \{\{1, 2, 3, 4\}, \{5, 6, 7\}, \{8, 9\}\}, \quad (6)$$

$$\frac{U}{R_2} = \{\{1, 2\}, \{3\}, \{4\}, \{5, 6, 7\}, \{8, 9\}\}. \quad (6)$$
Their positive areas about $X$ on $R_1$ and $R_2$ are $R_1(X) = R_2(X) = \{1, 2, 3, 4\}$. The weight of the knowledge characteristics $w_{R_1}(X) = \text{Card}(\text{POS}_{R_1}(X))/\text{Card}(U) = 4/9$, in which $\text{Card}(X)$ represents the number of the elements in the collection $X$. And the weight is also shown in $w_{R_2}(X) = \text{Card}(\text{POS}_{R_2}(X))/\text{Card}(U) = 4/9$. Thus $w_{R_1}(X) = w_{R_2}(X)$. It is obvious that the characteristics weights are the same, but the equivalence classes of these two characteristics are different.

In order to solve the problems above, we use the knowledge granularity to study the relationship between the various subsets in the complex sets of the equivalence classes and propose a method based on the knowledge granularity to compute the discernibility of knowledge characteristics. Then, the knowledge characteristics weights according to the relationship between the discernibility and the weights of knowledge characteristics will be determined.

3.1. The Discernibility of Knowledge Characteristics. We first give a definition about the discernibility of the knowledge characteristics.

Definition 1. Suppose that $K = (U, R)$ is a knowledge base, $R$ is the equivalence relation, and $r \in R$ is a characteristic. Let $U/R = \{X_1, X_2, \ldots, X_n\}$ and $U/(R - \{r\}) = \{Y_1, Y_2, \ldots, Y_m\}$. Then, the discernibility of $r$ is denoted by

$$
\text{Dis}(r) = \text{Dis}(R) - \text{Dis}(R - \{r\}).
$$

(7)

By Definition 1, we know that the larger $\text{Dis}(r)$ is, the more discernible the ability of $r$ becomes. When we select two objects randomly on $U$, there are $|U|^2$ ways. After adding characteristic $r$ into $(R - \{r\})$, the characteristic discernibility increases from $|R - \{r\}|$ to $|R|$. Thus, the number of equivalence classes is more than or equal to the original set. Thus, the ability of such discernibility is improved, and the discernibility increases.

Theorem 2. Let $r \in R, U/R = \{X_1, X_2, \ldots, X_n\}, U/(R - \{r\}) = \{Y_1, Y_2, \ldots, Y_m\}$, and denote $\text{Dis}(r)$ as discernibility of $r$; then there is $0 \leq \text{Dis}(r) \leq 1 - 1/|U|$.

Proof. From (4) and (5), we have

$$
\text{Dis}(r) = \text{Dis}(R) - \text{Dis}(R - \{r\})
= (1 - \text{GD}(R)) - (1 - \text{GD}(R - \{r\}))
= \sum_{j=1}^{m} \frac{|Y_j|^2}{|U|^2} - \sum_{i=1}^{n} \frac{|X_i|^2}{|U|^2}
= \frac{\left(\sum_{j=1}^{m} |Y_j|^2 - \sum_{i=1}^{n} |X_i|^2\right)}{|U|^2}.
$$

(8)

After adding characteristic $r$ into $(R - \{r\})$, the characteristic discernibility increases from $|R - \{r\}|$ to $|R|$, and the number of equivalence classes increases. Thus, there exists $Y_j \in U/(R - \{r\}) (1 \leq j \leq m)$ such that $Y_j = \bigcup_{k=1}^{n} X_k$. And we have

$$
|Y_j|^2 = \left(\frac{\sum_{k=1}^{n} |X_k|^2}{|U|^2}\right)^2,
$$

(9)

which shows $\text{Dis}(r) \geq 0$.

When the granularity of $[X]_R$ attains its minimum, there is only one element in $X_j$. When $U/(R - \{r\}) = \{Y_1, Y_2, \ldots, Y_m\}$ reaches the universe $U$, $\text{Dis}(r)$ reaches its maximum. Then we obtain

$$
\text{Dis}(r) = \frac{\left(\sum_{j=1}^{m} |Y_j|^2 - \sum_{i=1}^{n} |X_i|^2\right)}{|U|^2}
= \frac{(U^2 - |U|)}{|U|^2}
= 1 - \frac{1}{|U|}.
$$

(10)

Thus, $0 \leq \text{Dis}(r) \leq 1 - 1/|U|$ is proved.

3.2. Method to Determine Characteristics Weight. To propose our new characteristics weight method, we further give two definitions.

Definition 3. Suppose that $K = (U, R)$ is a knowledge base and $R = C \cap D$, where $C$ denotes the condition characteristics and $D$ denotes the decision characteristics. $[X]_R = U/D = \{X_1, X_2, \ldots, X_n\}$ identifies the equivalence classes on the universe $U$ equivalence partitioned by the decision characteristics $D$. $\text{Dis}(C)$ is the discernibility of $C$ on the universe $U$. The discernibility of the knowledge characteristics on $[X]_R$ is defined as

$$
\text{KCDis}(C) = \rho_c (X) \text{Dis}(C).
$$

(11)

According to (2) and (5), we have the following formulation of KCDis($C$):

$$
\text{KCDis}(C) = \rho_c (X) \left(1 - \text{GD}(C)\right)
= \frac{\left|R_c (X)\right|}{\left|R_c (X)\right| \left(1 - \frac{\left|R_c (X)\right|^2}{|U|^2}\right)}
= \frac{\left(\sum_{j=1}^{m} |Y_j|^2 \left(\left|R_c (X)\right|/\left|R_c (X)\right|^2\right)\right)}{|U|}.
$$

(12)

Definition 4. Suppose that $K = (U, R)$ is a knowledge base and $R = C \cap D$, where $C$ is the condition characteristics and $D$ is the decision characteristics. $[X]_R = U/D = \{X_1, X_2, \ldots, X_n\}$ identifies the equivalence classes on the universe $U$ equivalence partitioned by the decision characteristics $D$. For condition characteristics $c \in C$, the
discernibility of $C$ is $\text{KCD}_{\text{Dis}}(C)$ and the discernibility of $(C - \{c\})$ is $\text{KCD}_{\text{Dis}}(C - \{c\})$. Then the discernibility of the $c$ $(\forall c \in C)$ is defined as

$$\text{KCD}_{\text{Dis}}(c) = \text{KCD}_{\text{Dis}}(C) - \text{KCD}_{\text{Dis}}(C - \{c\}). \quad (13)$$

According to Definitions 3 and 4, we present a new formula to compute the weight of characteristic in the following definition. Detailed computation process is shown in Algorithm 1.

**Definition 5.** Suppose that $K = (U, R)$ is a knowledge base and $R = C \cap D$, where $C$ denotes the condition characteristics and $D$ denotes the decision characteristics. $[X]_R = U / D = \{X_1, X_2, \ldots, X_n\}$ identifies the equivalence classes on the universe $U$ equivalence partitioned by the decision characteristics $D$. $\text{KCD}_{\text{Dis}}(C)$ is the discernibility of the knowledge characteristics on $[X]_R$ equivalence partitioned by the condition characteristics $C$. For any conditional characteristics $c \in C$, the weight of the characteristic is defined as

$$W(c) = \frac{\text{KCD}_{\text{Dis}}(c)}{\sum_{c \in C} \text{KCD}_{\text{Dis}}(c)}. \quad (14)$$

**Theorem 6.** Assume that $X = U / D = \{X_1, X_2, \ldots, X_n\}$ is the equivalence class on the universe $U$ equivalence partitioned by the characteristics $D$. For any condition characteristics $c \in C$, $\text{KCD}_{\text{Dis}}(c)$ is the discernibility of $c$ to $U$, and it satisfies $0 \leq \text{KCD}_{\text{Dis}}(c) \leq 1 - 1/|U|$.

**Proof.** By the rough set theory, we know that

$$0 \leq \rho_c(X) = \frac{\text{Dis}(C)}{\text{Dis}(C) + \text{Dis}(C - \{c\})} \leq 1. \quad (15)$$

According to Theorem 2, we have $0 \leq \text{Dis}(C) \leq 1 - 1/|U|$. Thus it is easy to check that $0 \leq \text{KCD}_{\text{Dis}}(c) = \rho_c(X)\text{Dis}(C) \leq 1 - 1/|U|$.

**Theorem 7.** Assume that $X = U / D = \{X_1, X_2, \ldots, X_n\}$ is the equivalence class on the universe $U$ equivalence partitioned by the decision characteristics $D$. For any condition characteristics $c \in C$, $\text{KCD}_{\text{Dis}}(c)$ is the discernibility of $c$ to $U$. Then

(1) if $P$ and $Q$ are two equivalence relations on $U$ and $P \subseteq Q$, then $\text{KCD}_{\text{Dis}}(P) \leq \text{KCD}_{\text{Dis}}(Q)$;

(2) $0 \leq \text{KCD}_{\text{Dis}}(c) \leq 1 - 1/|U|$.

**Proof.** According to Definition 3, there are $Y = \text{R}_{c}(X) = \{Y_1, Y_2, \ldots, Y_m\}$ and $Z = \text{R}_{C-\{c\}}(X) = \{Z_1, Z_2, \ldots, Z_m\}$. According to (11), there is

$$\text{KCD}_{\text{Dis}}(c) = \rho_c(X)\text{Dis}(C) - \rho_{c - \{c\}}(X)\text{Dis}(C - \{c\})$$

$$= \frac{\sum_{i=1}^n |X_i| \left( |R_{c}(X_i)| / |R_{C}(X_i)| \right) \left( 1 - |R_{c}(X_i)| / |U| \right) - \left( |R_{C-\{c\}}(X_i)| / |R_{C-\{c\}}(X_i)| \right)\left( 1 - |R_{C-\{c\}}(X_i)| / |U| \right)}{|U|}$$

$$= \frac{\sum_{i=1}^n |X_i| \left( |R_{c}(X_i)| / |R_{C}(X_i)| \right) \sum_{j=1}^n \left( 1 - |Y_j| / |U| \right) - \left( |R_{C-\{c\}}(X_i)| / |R_{C-\{c\}}(X_i)| \right)\sum_{k=1}^m \left( 1 - |Z_k| / |U| \right)}{|U|}$$

(16)
(1) For the universe $U$, $P$ and $Q$ are two equivalence relations on the universe $U$. Let $P = Q - \{q\}$ ($\forall q \in Q$). There are $Y = \mathcal{R}_Q(X) = \{Y_1, Y_2, \ldots, Y_n\}$ and $Z = \mathcal{R}_{Q-\{q\}}(X) = \{Z_1, Z_2, \ldots, Z_m\}$. There exists $Z_j \in X/(Q - \{q\})$ ($1 \leq j \leq m$) such that $Z_j = \bigcup_{k=1}^{r} Y_k$.

For the universe $U$, there are $\mathcal{R}_Q(X) \supseteq \mathcal{R}_{Q-\{q\}}(X) = \mathcal{R}_P(X)$. So the following is satisfied:

$$\rho_P(X) = \left| \frac{\mathcal{R}_P(X)}{\mathcal{R}_P(X)} \right| \leq \rho_Q(X) = \left| \frac{\mathcal{R}_Q(X)}{\mathcal{R}_Q(X)} \right|. \quad (17)$$

When $Z_j = \bigcup_{k=1}^{r} Y_k$, we have $|Z_j|^2 = \sum_{k=1}^{r} |Y_k|^2 = (\sum_{k=1}^{n} |Y_k|^2) \geq \sum_{k=1}^{n} |Y_k|^2$ and

$$\sum_{j=1}^{\infty} \left( 1 - \frac{|Y_j|^2}{|U|^2} \right) \geq \sum_{k=1}^{n} \left( 1 - \frac{|Z_k|^2}{|U|^2} \right). \quad (18)$$

Substituting (18) into (16), we obtain

$$\sum_{j=1}^{\infty} \left| X_j \right| \left( \frac{\mathcal{R}_Q(X_j)}{|U|^2} \right) \geq \sum_{j=1}^{\infty} \left| X_j \right| \left( \frac{\mathcal{R}_P(X_j)}{|U|^2} \right) \geq \sum_{j=1}^{\infty} \left| X_j \right| \left( \frac{\mathcal{R}_P(X_j)}{|U|^2} \right) \geq \sum_{k=1}^{\infty} \left| Z_k \right|. \quad (19)$$

Therefore, KCDis$(P) \leq$ KCDis$(Q)$.

(2) From (16) and (19), we have KCDis$(c) \geq 0$. When $C$ becomes the universe $U$, it partitions the universe $U$ into equivalence classes (one class comprises individual elements). For this case, KCDis$(c)$ reaches its maximum KCDis$(c) = 1/|U|$. Therefore, $0 \leq$ KCDis$(c) \leq 1 - 1/|U|$ is obtained.

4. Experimental Results

In this section, some experiments are used to show the effectiveness of our new method. The data used in our experiments come from the Pima Indians Diabetes Data Set, which includes a total of 768 cases, of which 392 are valid, and the rest of the data cases’ characteristics values are missing. Note that the Pima Indians Diabetes Data Set is no longer available due to permission restrictions.

In actual computations, we use these 392 cases for experimentation. The condition characteristics information includes “plasma glucose concentration at 2 hours in an oral glucose tolerance test”, “diastolic blood pressure (mm Hg)”, “triceps skin fold thickness (mm)”, “2-hour serum insulin (mu U/ml)”, “body mass index (weight in kg/(height in m)²)” . The data set is given in Table 1, where “c1”, “c2”, “c3”, “c4”, and “c5” denote the condition characteristics, respectively. “d” stands for the decision characteristics “class variable (0 or 1)”. Then the condition characteristics values are discretized to different levels as “A, B, C” or “A, B, C, D”; see Table 2.

<table>
<thead>
<tr>
<th>U</th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>89</td>
<td>66</td>
<td>23</td>
<td>94</td>
<td>28.1</td>
<td>0</td>
</tr>
<tr>
<td>U2</td>
<td>137</td>
<td>40</td>
<td>35</td>
<td>168</td>
<td>43.1</td>
<td>1</td>
</tr>
<tr>
<td>U3</td>
<td>78</td>
<td>50</td>
<td>32</td>
<td>88</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>U4</td>
<td>197</td>
<td>70</td>
<td>45</td>
<td>543</td>
<td>30.5</td>
<td>1</td>
</tr>
<tr>
<td>U5</td>
<td>189</td>
<td>60</td>
<td>23</td>
<td>846</td>
<td>30.1</td>
<td>1</td>
</tr>
<tr>
<td>U6</td>
<td>166</td>
<td>72</td>
<td>19</td>
<td>175</td>
<td>25.8</td>
<td>1</td>
</tr>
<tr>
<td>U7</td>
<td>118</td>
<td>84</td>
<td>47</td>
<td>230</td>
<td>45.8</td>
<td>1</td>
</tr>
<tr>
<td>U8</td>
<td>103</td>
<td>30</td>
<td>38</td>
<td>83</td>
<td>43.3</td>
<td>0</td>
</tr>
<tr>
<td>U9</td>
<td>115</td>
<td>70</td>
<td>30</td>
<td>96</td>
<td>34.6</td>
<td>1</td>
</tr>
<tr>
<td>U10</td>
<td>126</td>
<td>88</td>
<td>41</td>
<td>235</td>
<td>39.3</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>U384</td>
<td>100</td>
<td>84</td>
<td>33</td>
<td>105</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>U385</td>
<td>81</td>
<td>74</td>
<td>41</td>
<td>57</td>
<td>46.3</td>
<td>0</td>
</tr>
<tr>
<td>U386</td>
<td>187</td>
<td>70</td>
<td>22</td>
<td>200</td>
<td>36.4</td>
<td>1</td>
</tr>
<tr>
<td>U387</td>
<td>121</td>
<td>78</td>
<td>39</td>
<td>74</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>U388</td>
<td>181</td>
<td>88</td>
<td>44</td>
<td>510</td>
<td>43.3</td>
<td>1</td>
</tr>
<tr>
<td>U389</td>
<td>128</td>
<td>88</td>
<td>39</td>
<td>110</td>
<td>36.5</td>
<td>1</td>
</tr>
<tr>
<td>U390</td>
<td>88</td>
<td>58</td>
<td>26</td>
<td>16</td>
<td>28.4</td>
<td>0</td>
</tr>
<tr>
<td>U391</td>
<td>101</td>
<td>76</td>
<td>48</td>
<td>180</td>
<td>32.9</td>
<td>0</td>
</tr>
<tr>
<td>U392</td>
<td>121</td>
<td>72</td>
<td>23</td>
<td>112</td>
<td>26.2</td>
<td>0</td>
</tr>
</tbody>
</table>

According to Algorithm 1, the following characteristics weights can be obtained:

$$W(c_1) = 0.3625,$$
$$W(c_2) = 0.0451,$$
$$W(c_3) = 0.2388,$$
$$W(c_4) = 0.2848,$$
$$W(c_5) = 0.0688.$$

Two experiments are conducted to show the advantages of our new method. The first experiment is to compare different rough set based methods with our method. The second one is to compare the AHP and PCA methods with our method. Both comparisons can show that our new proposed method is more effective than those methods.

In the first experiment, we also choose two rough set based methods. One is based on the dependence in rough set theory to calculate the characteristics weight. The other is based on rough sets and conditional information entropy.

In knowledge bases $K = (U, R)$ and $R = C \cap D$, the dependence of the characteristic is defined as $\gamma_C(D) = |\text{POS}_R(D)|/|U|$. The characteristics importance $\text{Sig}(c) = \gamma_C(D) - \gamma_{C-\{c\}}(D)$. Then the characteristics weight is $W_i(c) = \text{Sig}(c)/\sum_{c \in C} \text{Sig}(c)$ [39]. By calculation, we have

$$W_i(c_1) = 0.5,$$
$$W_i(c_2) = 0,$$
$$W_i(c_3) = 0.3,$$
Table 2: The discretized Pima Indians Diabetes Data Set.

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>U2</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td>U3</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>U4</td>
<td>D</td>
<td>B</td>
<td>D</td>
<td>D</td>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>U5</td>
<td>D</td>
<td>B</td>
<td>B</td>
<td>D</td>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>U6</td>
<td>D</td>
<td>B</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>U7</td>
<td>B</td>
<td>A</td>
<td>D</td>
<td>C</td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td>U8</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td>U9</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>U10</td>
<td>C</td>
<td>C</td>
<td>D</td>
<td>C</td>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>U384</td>
<td>D</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>U385</td>
<td>A</td>
<td>B</td>
<td>D</td>
<td>A</td>
<td>C</td>
<td>1</td>
</tr>
<tr>
<td>U386</td>
<td>D</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>U387</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>0</td>
</tr>
<tr>
<td>U388</td>
<td>D</td>
<td>B</td>
<td>D</td>
<td>D</td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td>U389</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td>U390</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>U391</td>
<td>B</td>
<td>B</td>
<td>D</td>
<td>C</td>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>U392</td>
<td>C</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ W_1(c_4) = 0.2, \]
\[ W_1(c_5) = 0. \]  \hspace{1cm} (21)

In knowledge bases \( K = (U,R) \) and \( R = C \cap D \), the dependence of the characteristic is defined as \( I(D \mid C) = \sum_{m=1}^{m} \left( |C_j|/|U| \right) \sum_{j=1}^{n} \left( |D_j \cap C_j|/|C_j| \right) (1 - |D_j \cap C_j|/|C_j|) \), and the characteristics importance \( \text{Sig}(c) = I(D \mid C - \{c\}) - I(D \mid C) \), and the characteristics weight is \( W_2(c) = \text{Sig}(c)/\sum_{a \in C} \text{Sig}(a) \) [40]. By calculation, we have

\[ W_2(c_1) = 0.5, \]
\[ W_2(c_2) = 0, \]
\[ W_2(c_3) = 0.2857, \]
\[ W_2(c_4) = 0.2143, \]
\[ W_2(c_5) = 0. \]  \hspace{1cm} (22)

In Table 3, we list the weighting results of the three methods based on rough set. Figure 1 clearly shows their comparison. From Table 3 and Figure 1, it shows that when the methods based on the dependence of rough set and the method based on the rough set and conditional information entropy are used to calculate the characteristics weights, the weights of “c2” and “c5” are redundant. But when the proposed method is used to calculate the characteristics weights, the results were not redundant. There is a little relation between “diastolic blood pressure (mm Hg)”, “body mass index (weight in kg/(height in m)^2)”, and diabetes, but they are related. So, from this point of view, the new method is more accurate than the other two rough set-based methods.

In the second experiment, the AHP method and the PCA method are used to calculate the characteristics weight. We also compare their results with ours.

For the AHP method, we construct the analytic hierarchy matrix according to the opinion of medical experts [41]. Then we obtain the weights:

\[ W_3(c_1) = 0.0604, \]
\[ W_3(c_2) = 0.1012, \]
\[ W_3(c_3) = 0.3103, \]
\[ W_3(c_4) = 0.1815, \]
\[ W_3(c_5) = 0.3465. \]  \hspace{1cm} (23)

For the PCA method, we select the representative variables through the transformation of multiple variables. Then the SPSS software is used to seek the explanation of the total variance and component of the matrix. We take principal components variance contribution rate as weight [41] and finally normalize them to get the weights:

\[ W_4(c_1) = 0.432, \]
\[ W_4(c_2) = 0.1114, \]
\[ W_4(c_3) = 0.0978, \]
\[ W_4(c_4) = 0.2568, \]
\[ W_4(c_5) = 0.1019. \]  \hspace{1cm} (24)
### Table 3: Comparison of three methods based on rough set.

<table>
<thead>
<tr>
<th>Method</th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
</tr>
</thead>
<tbody>
<tr>
<td>The method based on the dependence of rough set</td>
<td>0.5000</td>
<td>0</td>
<td>0.3000</td>
<td>0.2000</td>
<td>0</td>
</tr>
<tr>
<td>The method based on rough set and conditional information entropy</td>
<td>0.5000</td>
<td>0</td>
<td>0.2857</td>
<td>0.2143</td>
<td>0</td>
</tr>
<tr>
<td>The method of this paper</td>
<td>0.3625</td>
<td>0.0451</td>
<td>0.2388</td>
<td>0.2848</td>
<td>0.0688</td>
</tr>
</tbody>
</table>

### Table 4: Comparison of different types of methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>c1</th>
<th>c2</th>
<th>c3</th>
<th>c4</th>
<th>c5</th>
</tr>
</thead>
<tbody>
<tr>
<td>The AHP method</td>
<td>0.0604</td>
<td>0.1012</td>
<td>0.3103</td>
<td>0.1815</td>
<td>0.3465</td>
</tr>
<tr>
<td>The PCA method</td>
<td>0.432</td>
<td>0.1114</td>
<td>0.0978</td>
<td>0.2568</td>
<td>0.1019</td>
</tr>
<tr>
<td>The method of this paper</td>
<td>0.3625</td>
<td>0.0451</td>
<td>0.2388</td>
<td>0.2848</td>
<td>0.0688</td>
</tr>
</tbody>
</table>

The weighting results are given in Table 4. Figure 2 shows the comparison between the proposed method and two well-known methods. From Table 4 and Figure 2, it is easy to check that the rank of the results calculated with our method is "c1" > "c4" > "c3" > "c5" > "c2". It shows that there is a closed relation between "plasma glucose concentration at 2 hours in an oral glucose tolerance test" and diabetes, and there is a little relation between "diastolic blood pressure (mm Hg)" and diabetes. These results are synthetic optimization on the results calculated by AHP and PCA from Figure 2. By consulting the medical experts, the results calculated by our method are more accordant with the actual situation.

However, the Analytical Hierarchy Process (AHP) method is based on the subjective judgment of the experts and the Principal Component Analysis (PCA) method needs to extract representative principal components and increase an additional a priori information and evaluation criteria. Therefore, these two methods cannot objectively reflect the weight distribution. The new method does not need the prior knowledge, but the obtained weights are in line with the actual situation.

From the above discussion, the weighting method based on rough set can avoid the arbitrariness of subjective judgment. In addition, the weighting method with granularity theory can effectively avoid taking nonredundant characteristics as redundant characteristics. We can conclude that our new method reasonably distributes the weight for each characteristic. The weights basically reflect the importance of each characteristic and can also objectively reflect the actual situation of the patient’s body. Thus, the proposed method is a powerful method in knowledge classification.

### 5. Conclusions

Knowledge characteristics can help us have a good understanding of the knowledge base. The determination of knowledge characteristics weight can help us effectively classify the knowledge base, so as to achieve the purpose of knowledge management and decision making. In this paper, based on rough set theory and knowledge granularity theory, the weights of knowledge characteristics are determined. Experimental results show that the proposed method can effectively avoid taking nonredundant characteristics as redundant characteristics and can effectively determine the weights of knowledge characteristics.

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 61472256, no. 61771265) and the Natural Science Foundation of Jiangsu Province (no. BK20151272).
References


