Research Article

Research on Grey Relational Clustering Model of Multiobjective Human Resources Based on Time Constraint

Xiaowen Wang and Yijun Mu

School of Economics, Lanzhou University, Lanzhou 730000, China

Correspondence should be addressed to Yijun Mu; muyj16@lzu.edu.cn

Received 2 March 2021; Revised 26 March 2021; Accepted 3 April 2021; Published 17 April 2021

Academic Editor: Shah Nazir

Copyright © 2021 Xiaowen Wang and Yijun Mu. 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.

The existing grey relational clustering method has limitations in the application of multidimensional sequences and cannot directly calculate the grey correlation degree between unequal-length sequences. In this paper, by introducing the multidimensional dynamic DTW distance into the existing 3D grey relational model, a new grey relational analysis model that can be applied to multidimensional data is proposed, which is based on DTW distance. The model does not require one-to-one correspondence of data points but evaluates the similarity of its geometric curves by calculating the shortest distance between sequences. In addition, since the traditional grey correlation clustering method is implemented, the method first extracts the reference sequence from the observation sequence and then calculates the similarity between the observation objects by calculating the grey correlation degree between each sequence and the reference sequence, so each object only needs to be calculated once. The experimental results show that the multidimensional grey correlation degree based on DTW distance and the grey relational clustering model oriented to multidimensional data are more accurate than other existing methods. Finally, the grey relational clustering method of multidimensional data is used to analyze the multiobjective human resource grey relational clustering model under time constraints, and the validity of the model is verified.

1. Introduction

Clustering, as an important data analysis method, has attracted the extensive attention of many scholars and is a research hotspot in machine learning, statistics, computer science, and other fields. Most of the existing clustering methods are aimed at the data with massive information [1]. For those data with a small sample size, insufficient information, and unclear sample rules, these methods often cannot get accurate results. The main object of grey relational clustering is such data [2]. It measures the similarity of the observation systems by a grey relational analysis model and simplifies the complex system by dividing similar objects into the same class. In recent years, many scholars have improved and optimized it and applied it to different fields. Grey relational analysis and its model were proposed by Professor Deng Julong in 1985 [3]. As an important research field in grey system theory, it mainly judges the similarity between two sequences according to the geometric shape of the sequence curve [4]. Linear interpolation is usually used to transform the observed data of the discrete behavior of the observation system into piecewise continuous lines, and then a corresponding model is constructed according to the geometric characteristics of the lines to judge the similarity between the sequences, including the characteristics of distance, area, and slope [5]. The more similar objects are observed, the more similar their geometric characteristics are. Because the model is not affected by the sample size and the rule of the sample distribution. In addition, the model also has the characteristics of small computation and convenient application [6].

Most of the existing grey clustering methods can be used to deal with one-dimensional data, so the research objects are limited [7]. To broaden the application scope of grey clustering, some scholars tried to combine grey theory with panel data and put forward the grey clustering method under...
panel data [8]. By combining the model with the traditional grey relational clustering method, good clustering results are achieved [9]. By using the principal component analysis method, the grey clustering coefficient matrix at different times is obtained by calculating the correlation degree between the score sequences of the principal components of two comprehensive factors [10]. However, the above method only considers the fluctuation of the indexes between adjacent objects, and if the panel data takes different time, redundant data often need to be supplemented or deleted, so the original data will be destroyed. In addition, some methods also have the problem of the small correlation between indexes in the class [10]. Aiming at these problems, some scholars also put forward a grey index correlation clustering model [11]. The model extracts the feature information from the original sequence by constructing the generating sequence and realizes the dimensionality reduction of the sequence [12]. The clustering rules of panel data can effectively avoid clustering sequences with small association degree into a group and can deal with unequal data. However, this method is easy to be affected by dimensionality reduction results, and the introduction of new uncertainties will affect the final clustering accuracy [13].

2. Basic Principles of Grey Relational Analysis and Cluster Analysis

2.1. Basic Principles of Grey Relational Analysis. We should analyze its components, distinguish the behavior factors and related factors, and make clear the characteristics of the system behavior; then, we should focus on the relationship between the system behavior and related factors; and finally, we should estimate and predict the system behavior [14]. When describing the system behavior and its related factors, that is, the feature mapping quantity of the system behavior and related factors, the feature mapping quantity is used to indirectly reflect the system behavior and related factors [15].

Definition 1. The sequence of behavioral characteristic data of behavioral factors $X_0 = (x_0(1), x_0(2), ..., x_0(n))$ is assumed. The sequence of behavioral characteristic data of related factors is as follows [16]:

$$ X_1 = (x_1(1), x_1(2), ..., x_1(n)), \\
X_j = (x_j(1), x_j(2), ..., x_j(n)), \\
X_m = (x_m(1), x_m(2), ..., x_m(n)). $$

(1)

The main calculation steps of Deng’s grey correlation are [17] as follows:

Step 1: order:

$$ X'_i = \frac{X_i}{x_i} = (x'_1(1), x'_2(2), ..., x'_n(n)), $$

$$ i = 0, 1, 2, ..., m. $$

(2)

Step 2: the difference sequence. Remember:

$$ \Delta_i(k) = |x'_i(k) - x'_j(k)|, $$

$$ \Delta_i = (\Delta_i(1), \Delta_i(2), ..., \Delta_i(n)), \quad i = 1, 2, ..., m. $$

(3)

Step 3: remember:

$$ M = \max_{i} \max_{k} \Delta_i(k), $$

$$ m = \min_{i} \min_{k} \Delta_i(k). $$

(4)

Step 4: correlation coefficient:

$$ \gamma_{0i}(k) = \frac{m + \xi M}{\Delta_i(k) + \xi M}, \quad \xi \in (0, 1), $$

$$ k = 1, 2, ..., n; i = 1, 2, ..., m. $$

(5)

Step 5: calculate relevance:

$$ \gamma_{0i} = \frac{1}{n} \sum_{k=1}^{n} \gamma_{0i}(k), \quad i = 1, 2, ..., m. $$

(6)

2.2. The Basic Principle of Grey Clustering Analysis Method

Definition 2. There are $n$ clustering objects and $m$ evaluation index values, and the $j$ index is divided into $s$ grey class, which is called the subclass of $j$ index [18].

The $s$ subclasses of each index have their own functions. Through these functions, different clustering objects are classified according to the values of the index. This function is called the whitening weight function [19]. A schematic diagram of a typical whitening weight function is shown in Figure 1.

The schematic diagram of whitening functions in four forms is shown in Figure 2.

(1) The typical whitening weight function can be described as a piecewise function:

$$ f^k_j(x) = \begin{cases} 
0, & x \notin [x^k_j(1), x^k_j(4)], \\
x - x^k_j(1), & x \in [x^k_j(1), x^k_j(2)], \\
x^k_j(2) - x^k_j(1), & x \in [x^k_j(2), x^k_j(3)], \\
1, & x \in [x^k_j(3), x^k_j(4)]. 
\end{cases} $$

(7)

(2) The lower bound measure whitening weight function can be described as a piecewise function:

$$ f^k_j(x) = \begin{cases} 
0, & x \notin [0, x^k_j(4)], \\
x, & x \in [0, x^k_j(3)], \\
x^k_j(4) - x, & x \in [x^k_j(3), x^k_j(4)]. 
\end{cases} $$

(8)
(3) Moderate measure whitening weight function can be described as a piecewise function:

\[
 f^k_j(x) = \begin{cases} 
 0, & x \notin [x^k_j(1), x^k_j(4)], \\
 \frac{x - x^k_j(1)}{x^k_j(2) - x^k_j(1)}, & x \in [x^k_j(1), x^k_j(2)], \\
 \frac{x^k_j(4) - x}{x^k_j(4) - x^k_j(2)}, & x \in [x^k_j(2), x^k_j(4)]. 
\end{cases}
\]

(4) The upper bound measure whitening weight functions can be described as piecewise function:

\[
 f^k_j(x) = \begin{cases} 
 0, & x < x^k_j(1), \\
 \frac{x - x^k_j(1)}{x^k_j(2) - x^k_j(1)}, & x \in [x^k_j(1), x^k_j(2)], \\
 1, & x \geq x^k_j(2). 
\end{cases}
\]
Panel data is an extended form of cross-sectional data in time dimension, which is composed of three dimensions: time, index, and object. The description of the panel data is shown in Table 1. Because panel data can better reflect the change of different attributes over time than cross-sectional data, panel data has more important research value [20].

3. Grey Relational Clustering Model for Multidimensional Data

3.1. Multidimensional DTW Distance and Its Multidimensional Grey Relational Analysis Model. Some scholars have extended the application of DTW distance to measure the similarity of multidimensional time series. There are two main research directions of multidimensional DTW. One is to convert multidimensional sequences into one-dimensional sequences and calculate the DTW distances between sequences. If the traditional dynamic time warping distance is used to process multidimensional sequences, the simplest way is to apply it to each dimension separately [19, 20]. The distance matrix is constructed by calculating the sum of the distances of each dimension of the sequence, and then the shortest curved path between the sequences is obtained from the distance matrix to match it. Before calculating the DTW distance between sequences, it is necessary to normalize the points of each dimension in the sequence, such as calculating its zero mean or variance [21].

**Definition 3.** Multidimensional dynamic time warping distance.

Given two sequences \( t \in R^{K \times L} \), and \( r \in R^{K \times L} \), assuming \( K \) is the number of attributes or dimensions of a data, \( tL \) and \( rL \) are the sequence lengths of the data sequences \( t \) and \( r \), respectively. The first step is to preprocess the sequence, which is to calculate the zero mean or unit variance of each dimension of the sequence. The second step is to use the following formula to calculate the distance matrix:

\[
d(i, j) = \sum_{k=1}^{K} |t(k,i) - r(k,j)|.
\]

Using the above formula, the vector normal distance between the points of the sequence is calculated and the corresponding distance matrix is constructed. Then, the shortest path between the sequences is obtained from the matrix. The DTW distance between sequences \( t \) and \( r \) is

\[
D(i, j) = d(i, j) + \min \left\{ D(i - 1, j), D(i, j - 1), D(i - 1, j - 1) \right\}.
\]

DTW distance considers a sequence as a point in m-dimensional space and then extracts a reference sequence from the original sequence so that the correlation between the two sequences depends not only on themselves but also on the sequence of other factors [21]. The three-dimensional grey relational degree model obtains the environmental parameters between two sequences by calculating the Minh space distance of the corresponding time between two sequences. However, this leads to the value of environmental parameters sometimes greater than the distance between the two sequences, making the grey correlation degree greater than 1. At this point, the model cannot satisfy the four axioms of relevance. The reference sequences are selected as the maximum and minimum differences of the two poles of each sequence based on the multidimensional grey correlation degree of DTW distance. Among them, the minimum difference between the two poles is always less than the shortest distance between the two sequences, which guarantees the normality of the grey correlation degree. Finally, by introducing multidimensional DTW distance into Deng’s grey relational degree calculation formula, a multidimensional grey relational analysis method based on DTW distance is proposed. The improvement idea of this method is shown in Figure 3. The figures contain three-dimensional grey relational grade on one side and multidimensional grey relational grades based on dynamic time warping distance on the other side [22].

The feasibility and validity of the multidimensional grey correlation degree based on DTW distance are verified by the corresponding grey correlation analysis of the data. The experimental procedures are as follows:

1. Step 1: get the maximum value of each dimension of each time, divide the value of the corresponding time of each observation factor by this value, and then get the corresponding initial value image.
2. Step 2: define parameter values.
3. Step 3: calculate the extremum.
4. Step 4: calculate the mingling distance and the multidimensional DTW distance between the observed objects and the reference sequences.
5. Step 5: calculate the three-dimensional grey correlation degree of the observation objects and the multidimensional grey correlation degree based on DTW distance, respectively.
6. Step 6: use the result of grey correlation degree to rank or cluster the observed objects.

The results of the multidimensional grey relational analysis model based on DTW distance are accurate and consistent when dealing with equal-length and unequal-length sequences. When the three-dimensional grey relational analysis model deals with unequal-length sequences, the results of correlation degree obtained by different data pretreatment methods are also different. When the missing data is filled with 0, the result of the correlation is relatively large. When the sequence is filled with the mean of adjacent data, because the new sequence obtained at this time is similar to the original sequence, the result is more accurate in economic benefit evaluation. However, when the model is clustered in five cities of Hunan Province, the result of the correlation is inaccurate. Three-dimensional grey relational grade and multidimensional grey relational grade calculation based on DTW distance are shown in Table 2.
The correlation coefficient of each sequence and ideal scheme is calculated. Constructing ideal scheme is calculated. Calculate the mingling distance between each sequence and ideal scheme. The correlation coefficient $R_1$ of each sequence and ideal scheme is calculated. The final grey relational grade is calculated: $R = R_1 / (R_1 + R_2)$.

<table>
<thead>
<tr>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1, \ldots, u_H$</td>
<td>$u_1, \ldots, u_H$</td>
<td>$u_1, \ldots, u_H$</td>
</tr>
<tr>
<td>$x_{i1}(t_1), \ldots, x_{iH}(t_1), x_{im}(t_1)$</td>
<td>$x_{i1}(t_2), \ldots, x_{iH}(t_2), x_{im}(t_2)$</td>
<td>$x_{i1}(t_T), \ldots, x_{iH}(t_T), x_{im}(t_T)$</td>
</tr>
<tr>
<td>$S_1$</td>
<td>$S_2$</td>
<td>$S_n$</td>
</tr>
</tbody>
</table>

**Table 1**: Panel data description form.

**Table 2**: Three-dimensional grey relational grade and multidimensional grey relational grade calculation based on DTW distance.

<table>
<thead>
<tr>
<th>Grey relational model</th>
<th>Raw data set</th>
<th>Missing data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multidimensional grey relational grade based on DTW distance</td>
<td>$S_2 &gt; S_3 &gt;&gt; S_4$</td>
<td>$S_2 &gt; S_3 &gt;&gt; S_4$</td>
</tr>
</tbody>
</table>
| Three-dimensional grey relational grade | $S_2 > S_3 > S_5 > S_4$ | Missing data complement mean; $S_2 > S_1 > S_5 > S_4 > S_3$ missing data complement mean $S_2 > S_3 >> S_4$

**3.2. Grey Relational Clustering for Multidimensional Data.**

To realize the grey relational clustering of multidimensional data, a grey relational clustering method for multidimensional data is proposed based on the idea of extracting reference sequences from original sequences by referring to the three-dimensional grey relational analysis method. Using this method, each sequence only needs to calculate the grey correlation degree once, instead of comparing two or more sequences like the traditional grey relational clustering method. Therefore, the calculation process of the method is more convenient and fast. In addition, this method does not require the same length of multidimensional sequence data, compared with the traditional grey relational clustering method, and its application scope is wider. The flow chart of the multidimensional grey relational clustering method is shown in Figure 4.

**Definition 4.** Grey relational clustering of multidimensional data.

Given the sequence group $X_1, X_2, \ldots, X_m$, suppose any sequence can be represented as $X_i = \{x_{i1}^1(1), \ldots, x_{in}^i(n)\}$, $\{x_{i1}^2(1), \ldots, x_{in}^i(n)\}, \ldots, \{x_{i1}^m(1), \ldots, x_{in}^j(n)\}$, where $x_i(j)$ represents the $j$th data of dimension $i$. The grey correlation degree between each sequence is calculated. According to these correlation values, multiple critical intervals are manually determined in interval $[0, 1]$. The flow chart of the multidimensional grey relational clustering method is shown in Figure 4.

Because the method does not need to preprocess the sequence to obtain the same length sequence as the traditional grey relational clustering method, it will not introduce new uncertainties into the data, so it can get better clustering results. In addition, because the traditional grey relational
clustering method needs to compare all objects in pairs when clustering, when there are more data objects, the amount of calculation will be very large. However, this clustering method does not need two-to-two comparisons between sequences, each object only needs to be computed once, and the calculation process is more convenient.

4. Experimental Analysis and Application

4.1. Experimental Analysis of Grey Relational Clustering Model for Multidimensional Data. Because the traditional grey relational clustering method needs to compare two sequences, its time complexity is high. To evaluate the accuracy of the grey relational clustering method for multidimensional data, the clustering experiments of Iris and Wine data sets in UCI data set were carried out, and the clustering effects of the two methods were evaluated by Rand Index [22].

The corresponding experimental steps are as follows:

(1) Step 1: get the maximum value of each time dimension of the sequence, divide the value of the corresponding time of each sequence by this value, and then get the corresponding initial value image.

(2) Step 2: the grey relational clustering method of multidimensional data uses the initial value image and selects the minimum or maximum values of each dimension at the corresponding time to form a reference sequence according to the characteristics of the data set.

(3) Step 3: calculate the initial sequence values separately.

(4) Step 4: calculate the mingling distance and the multidimensional dynamic time warping distance between each sequence and the reference sequence.

(5) Step 5: according to the third and fourth steps, the multidimensional grey scale correlation of DTW is calculated.

(6) Step 6: set the threshold interval, and the sequences in the same threshold interval are divided into the same category.

Grey relational clustering and comparison of traditional methods for multidimensional data are shown in Table 3. Grey relational clustering of multidimensional data and Rand Index value of traditional grey relational clustering are shown in Figure 5.

As can be seen from Table 3, the grey correlation clustering of multidimensional data is improved by nearly 30% compared with the traditional grey correlation clustering, and the clustering result is more accurate. In addition, the traditional grey relational clustering method needs two-to-two comparisons of sequences, and the grey
relational clustering method of multidimensional data only needs to calculate the grey relational degree for each object in clustering, so the calculation process is more convenient.

4.2. Application of Grey Relational Clustering Method for Multidimensional Data. In this paper, the UCI data set “individual household electric power consumption” is used to design the experiment, and the grey relational clustering method for multidimensional data is used to realize the clustering of similar months of electricity consumption, and the household’s electricity consumption law for four years is analyzed.

The experiment takes the monthly electricity consumption as the object of observation with 12 objects. However, because the original data has more than 2 million, and the grey system research object is “small sample, poor information” uncertain data. Therefore, we need to pre-process the data and extend the sampling interval from one minute to one day. Therefore, the sequence length of each object is the number of days observed by electricity. However, the number of days per month is different, resulting in different lengths of data for each object in the sequence. Because the three-dimensional grey relational model needs to calculate Minh’s distance between sequences, it cannot deal with sequences of different lengths. Therefore, we need to delete long sequences before using this method. Multidimensional grey relational grade based on DTW distance is not needed.

Three-dimensional grey relational analysis model and its multidimensional grey relational analysis model clustering results based on DTW distance are shown in Table 4. The comprehensive score of fiscal input and output factors of science and technology is shown in Figure 6.

To more intuitively reflect the content of Table 4, the final clustering results are displayed in Figure 6. From the analysis of household electricity consumption based on the multidimensional grey relational analysis model of DTW distance, the clustering results in each month of the four years have little difference, but the clustering results obtained by the three-dimensional grey relational analysis model have great changes.

In view of the limitation of the existing grey relational clustering methods in the application of multidimensional sequences and not being able to directly calculate the grey relational degree between unequal-length sequences, in this paper, by introducing the multidimensional dynamic DTW distance into the existing three-dimensional grey relational model, a new grey relational analysis model, multidimensional grey relational degree based on DTW distance, is proposed, which can be applied to multidimensional data. The model does not need data point-to-point correspondence but evaluates the similarity of geometric curves by calculating the shortest distance between sequences. Finally, the grey relational clustering method of multidimensional data is used to analyze the multiobjective grey relational clustering model of human resources under time constraints, and the validity of the model is verified.

<table>
<thead>
<tr>
<th>Data set name</th>
<th>Multidimensional grey relational clustering</th>
<th>Traditional grey relational clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris data set</td>
<td>0.9978</td>
<td>0.6983</td>
</tr>
<tr>
<td>Wine data set</td>
<td>0.9598</td>
<td>0.6434</td>
</tr>
</tbody>
</table>

Figure 5: Grey relational clustering of multidimensional data and Rand Index value of traditional grey relational clustering.
5. Conclusion

The existing models still have limitations in the application of uncertain data and multidimensional data. As the data is collected in real life, there will be a possibility of data missing, and the existing grey relational analysis model requires that the sequence length is equal in application. If the length of the sequence is not equal, the missing data need to be filled or long sequences deleted to preprocess data, which will increase the uncertainty of the original sequence, so the accuracy of the grey relational degree is not enough, thus affecting the accuracy of the final clustering results. In addition, the characteristics of many objects need to collect multiple dimensional data to reflect their characteristics. However, most of the existing grey relational analysis models are based on one-dimensional data. By introducing the multidimensional DTW distance into the three-dimensional grey relational analysis model, the problem that the three-dimensional grey relational analysis model cannot deal with the uncertain sequence directly is solved. Compared with the traditional three-dimensional grey relational analysis model, the experimental results show that the accuracy of this method is higher than that of the three-dimensional grey relational analysis model in dealing with unequal and equal-length sequences. On the basis of multidimensional grey relational degree based on DTW distance and traditional grey relational clustering method, a new grey relational clustering method is proposed. By improving the traditional grey relational clustering method, this method can be applied to multidimensional sequence clustering. The effectiveness of the proposed approach is demonstrated through experimental results.

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.

References


