International Journal of Database Theory and Application
Vol.9, No.4 (2016), pp. 187-194
http://dx.doi.org/10.14257/ijdta.2016.9.4.17

A Multi-Index Grey Relational Data Mining Model of Complex System Based on Grey System Theory

Xiaohui Wang and Jianfen Liu

Collage of Computer Science and Technology, Pingdingshan University, Pingdingshan, Henan, China
Lsp1519@163.com

Abstract

Data mining is a complex systematic engineering. For complex systems that is influenced by multi indexes, the relevance, hierarchy and fuzziness among indexes pose great challenges to data mining of complex system. Therefore, the paper proposes a multi-index grey relational data mining model of complex system based on grey system theory. First, the paper constructs index sets and grey categories for data mining objectives, and, under the different grey categories, builds up grey classic domain and grey partial domain for the aforementioned index sets. Next, normalized processing is undertaken to unify scales of different kinds of insets, whose weight generating algorithm is also provided herein. The paper then establishes a grey relational coefficient model and a grey relational degree model whose value is used to obtain the grey categories of the data mining objectives. Finally, the paper verifies and expounds multi-index grey relational data mining model by applying it to real-life practice. The results show the effectiveness and feasibility of this model.

Keywords: *grey system theory; grey relation analysis; complex system; data mining; model and algorithm*

1. Introduction

Data mining is a popular study in the field of artificial intelligence. Generally, with algorithm, latent information from massive data is searched for reuse. This process enriches knowledge that is required for design, thus enhancing the efficiency and quality of design. For this, numerous specialists and scholars have undertaken a string of research and discussions and obtains a series of achievements in application fields. However, as they have different perspectives in solving problems, present data mining models and approaches tend to be applied with specific limitations, such as massive data, strong logical reasoning, or establishment of precise mining models. For this reason, the paper researches on multi-index data mining in complex system. By analyzing its feature, and based on grey system theory [8-10], the paper proposes an improved multi-index grey relational data mining model in complex system. This model provides a new approach to addressing issues in multi-type and multi-index data mining in large complex system.

2. Concepts of Grey Relational Theory

Grey system theory, as a discipline of applied mathematics, studies uncertain systems in which part of the information is known and part of the information is unknown. The system that it applies to is small in samples and poor in information, and has clear extension and unclear connotation. According to this theory, levels of recognition, information and decision determine whether the information is complete or not. There is connection between higher-level systems and lower-level systems with regard to the amounts of determinate information. Known information provides a sufficient way to

ISSN: 2005-4270 IJDTA

unveil rules of systems. As one of the main research approaches, grey relational analysis deals with problems of the above system. By establishing random and irregular multi-index grey relational analysis model in complex system, it determines relations among various random and irregular indexes, based on which it further analyzes, anticipates and digs situation data of developing and changing complex system. The grey system theory is widely applied in the field of engineering.

$$S(i)=\{s_1(i),\cdots,s_j(i),\cdots,s_n(i)\} \tag{1}$$

Accordingly, the time series of n indexes to data mining objectives in the complex system S is called as grey reference sequence $S(o)$, then

$$S(o)=\{s_1(o),\cdots,s_j(o),\cdots,s_n(o)\} \tag{2}$$

The grey relational coefficient ρ_{ij} of the j th index to $S(i)$ and $S(o)$ is as follows:

$$\rho_{ij}=\frac{\min_i\min_j\left|s_j(i)-s_j(o)\right|+\beta\max_t\max_k\left|s_j(i)-s_j(o)\right|}{\left|s_j(i)-s_j(o)\right|+\beta\max_t\max_k\left|s_j(i)-s_j(o)\right|} \tag{3}$$

Where the discriminant coefficient $\beta\in[0,1]$. It is generally taken as 0.5.

Grey relational grade between grey relational sequence $S(i)$ and grey reference sequence $S(o)$ is expressed as follows:

$$\delta_i=\frac{1}{n}\sum_{j=1}^{n}\rho_{ij} \tag{4}$$

In particular, when different indexes have different weights, and the weight of j is w_j, then equation (4) can be expressed as:

$$\delta_i=\sum_{j=1}^{n}\left(w_j*\rho_{ij}\right) \tag{5}$$

Where $\sum_{j=1}^{n}w_{kj}=1$.

For the situation mining of developing and changing complex system, given that the influential indexes are uncertain, and that the obtained data is obviously grey in possible disorder or in its representation way, grey relational analysis is well applicable to complex system with regard to digging and utilizing its latent intrinsic rules.

3. Multi-Index Grey Relational Data Mining Model of Complex System

3.1. Indexes Selection and Index Sets Establishment

To effectively undertake data mining of complex system, it is necessary to select in To effectively undertake data mining of complex system, it is necessary to select indexes that relates to the mining objectives in a reasonable way. Guided by the principle of scientificity, objectivity, comprehensiveness and measurability, the paper obtains indexes with both grey cluster analysis and comprehensive evaluation. Assume that there is n indexes that relates to the mining objectives, then the index set R is constructed as:

$$R = \{r_1, \cdots, r_j, \cdots, r_n\} \tag{6}$$

Obviously, with establishment of the index set R, the grey relational sequence $R(t)$ for any time t is expressed as follows:

$$R(t) = \{r_1(t), \cdots, r_j(t), \cdots, r_n(t)\} \tag{7}$$

3.2. Grey Categories and Grey Classical Domain

In order to identify precisely where the objective belongs, and with establishment of index set R, indexes should be specified for their grey categories. The partition size of grey categories depends on actual situations in data mining. Oversize partition means a broad scope of mined objective. Skimpy partition may results in lack of data support in categorizing objectives, which will neutralize efficacy. Therefore it is necessary to choose grey category partitions in a reasonable way. Assume that there are M grey categories of the index set R, the grey classical domain u_{kj} of index j under the grey category k is expressed as follows:

$$u_{kj} = \left[u_{kj}^{lef}, u_{kj}^{rig}\right], u_{kj}^{let} \le u_{kj}^{rig} \tag{8}$$

Correspondingly, the grey partial domain u_{oj} of index j under grey category k is expressed as follows:

$$u_{oj} = \left[u_{oj}^{let}, u_{oj}^{rig}\right] = \left[\min_{1\le k\le M} u_{kj}^{let}, \max_{1\le k\le M} u_{kj}^{rig}\right] \tag{9}$$

3.3 Normalized Treatment of Grey Numbers

The diversity of indexes determines that normalized treatment should be undertaken towards different types and dimensions of indexes so as to obtain their unified scales. The forms of treatment is generalized into the following forms:

For positive index j, if its grey classical domain $u_{kj} = \left[u_{kj}^{lef}, u_{kj}^{rig}\right]$, then after normalization, $v_{kj} = \left[v_{kj}^{lef}, v_{ki}^{rig}\right]$ turns into

$$v_{kj} = \left[\frac{u_{kj}^{lef} - u_{oj}^{lef}}{u_{oj}^{rig} - u_{oj}^{lef}}, \frac{u_{kj}^{rig} - u_{oj}^{lef}}{u_{oj}^{rig} - u_{oj}^{lef}}\right] \tag{10}$$

For negative index j, if its grey classical domain $u_{kj} = \left[u_{kj}^{lef}, u_{kj}^{rig}\right]$, then after normalization, $v_{kj} = \left[v_{kj}^{lef}, v_{ki}^{rig}\right]$ turns into

$$v_{kj} = \left[\frac{u_{oj}^{lef} - u_{kj}^{rig}}{u_{oj}^{rig} - u_{oj}^{lef}}, \frac{u_{oj}^{lef} - u_{kj}^{lef}}{u_{oj}^{rig} - u_{oj}^{lef}}\right] \tag{11}$$

For reaching index j, if its grey classical domain $u_{kj} = \left[u_{kj}^{lef}, u_{kj}^{rig}\right]$, and if the reaching optimal value is u_0, then after normalization, $v_{kj} = \left[v_{kj}^{lef}, v_{ki}^{rig}\right]$ turns into

$$v_{kj}=\left[1-\frac{u_{kj}^{let}-u_0}{\max\limits_{1\le k\le M}\left|u_{kj}^{let}-u_0\right|},1-\frac{u_{kj}^{rig}-u_0}{\max\limits_{1\le k\le M}\left|u_{kj}^{rig}-u_0\right|}\right] \tag{12}$$

In particular, if it reaches the optimal value $u_0=\left[u_0^{lef},u_0^{rig}\right]$, then after normalization, the grey classical domain $v_{kj}=\left[v_{kj}^{lef},v_{ki}^{rig}\right]$ can be divided into the following conditions:

(1) If $v_{kj}^p\le u_0^{lef}$, where $p=a$ or $p=b$, then

$$v_{kj}^p=1-\frac{u_0^{lef}-u_{kj}^p}{\max\left\{u_0^{lef}-u_{oj}^{lef},u_{oj}^{rig}-u_0^{rig}\right\}} \tag{13}$$

if $v_{kj}^p\in\left[u_0^{lef},u_0^{rig}\right]$, where $p=a$ or $p=b$, then

$$v_{lj}^p=1 \tag{14}$$

if $v_{kj}^p\le u_0^{rig}$, where $p=a$ or $p=b$, then

$$v_{kj}^p=1-\frac{u_{kj}^p-u_0^{rig}}{\max\left\{u_0^{lef}-u_{oj}^{lef},u_{oj}^{rig}-u_0^{rig}\right\}} \tag{15}$$

After the above normalized treatment, all of the classical domains and partial domains of influential factors are marked between 0 and 1, and satisfy the condition of $0\le v_{lj}^a\le v_{lj}^b\le 1$ and $0\le v_{oj}^a\le v_{0j}^b\le 1$. This thus guarantees uniformity of anticipation and analysis, and makes it more accurate and reliable for the results of multi-factor anticipating of complex system.

3.4 Grey Relational Data Mining Analysis

The paper obtains data of the index j and undertakes normalization treatment on it to acquire the index's magnitude v_j. Then the grey relational distance $D_{kj}^{\otimes}$ between the index j and the grey classical domain u_{kj} is expressed as follows:

$$D_{kj}^{\otimes}=\sqrt{\left|v_j-u_{kj}^{lef}\right|^2+\left|v_j-u_{kj}^{rig}\right|^2}/\sqrt{2} \tag{16}$$

And the grey relational coefficient $\zeta_{kj}^{\otimes}$ between the index j and the grey classical domain u_{kj} is expressed as follows:

$$\zeta_{kj}^{\otimes}=\frac{\min\limits_k\min\limits_j\left(D_{kj}^{\otimes}\right)+\beta\max\limits_k\max\limits_j\left(D_{kj}^{\otimes}\right)}{\left(D_{kj}^{\otimes}\right)+\beta\max\limits_k\max\limits_j\left(D_{kj}^{\otimes}\right)} \tag{17}$$

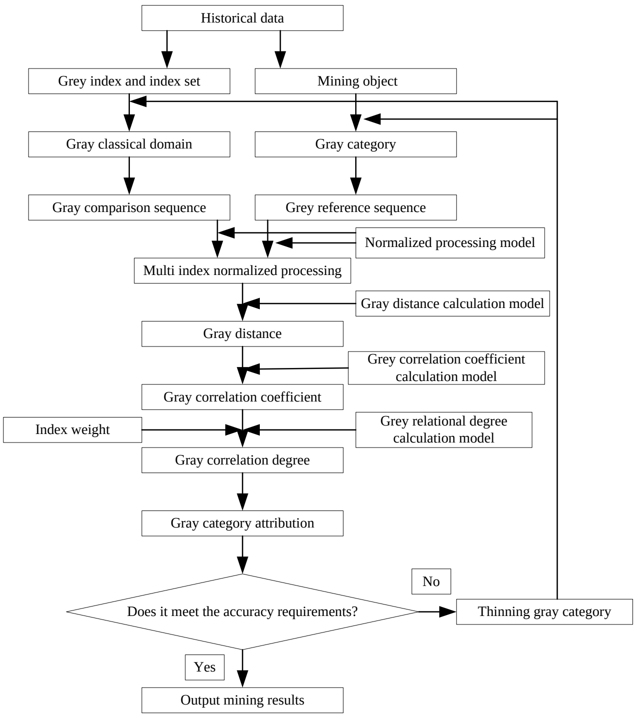
Given the weight w_j of the index j, the grey relational grade $\tau_k^{\otimes}$ between the mining objective and the grey category k is expressed as follows:

$$\tau_k^{\otimes}=\sum_{j=1}^{n}\left(w_j*\zeta_{kj}^{\otimes}\right) \tag{18}$$

According to grey relational degree $\tau_k^{\otimes}$, the paper determines which grey category the mining objective belongs to. The classic domain of the above category is the right magnitude zone of the mining objectives. It can be seen that the partition level of classic domain determines the preciseness degree of data mining. For this reason, if the mining objective fails to meet requirements on preciseness, process of partition treatment can be done repeatedly to indexes until the preciseness of the mining objective is enough.

3.5 Grey Modelling

As far as what has been mentioned, the basic procedure of the model herein is shown in Diagram 1 as follows:



4. Modelling Verification

Profits of a large corporation are closely related to its phased market plans of a premium brand. It is of great importance to dig and analyze phased revenue of the corporation. Therefore, the paper expounds the paper obtains the growth rate of products and profits of the premium brand in different marketing phases with investigation and analysis, which is shown in Table 1.

Table 1. Growth Rate of Products and Profits

Marketing phases	Growth rate			
	Product A	Product B	Product C	Profits
1	0.011	0.143	0.086	0.042
2	0.058	0.182	0.175	0.091
3	0.210	0.076	0.153	0.115
4	0.157	0.182	0.211	0.267
5	0.096	0.228	0.201	0.169
6	0.165	0.249	0.380	0.224
7	0.300	0.271	0.298	0.285
8	0.346	0.287	0.352	0.368
9	0.279	0.310	0.322	0.325

10	0.285	0.300	0.260	pending

Based on the growth rate of mining objectives, the classical domains of the obtained four grey categories for products A, B, and C, which are regarded as the relational indexes of data mining, is shown in Table 2.

Table 2. Classical Domains of Grey Categories

Grey categories	Mining range	classical domains		
		Index A	Index B	Index C
1	0-0.10	0.01-0.06	0.14-0.19	0.08-0.18
2	0.10-0.20	0.09-0.21	0.07-0.23	0.15-0.21
3	0.20-0.30	0.15-0.30	0.18-0.28	0.21-0.38
4	0.30-0.40	0.27-0.35	0.28-0.31	0.32-0.36

Based on the given grey distance and grey relational coefficient computing model herein, the paper obtains corresponding computing results, which is shown in Table 3.

Table 3. Grey Distances and Grey Relational Coefficient

Grey categories	Index A		Index B		Index C	
	distance	Relational categories	distance	Relational categories	distance	Relational categories
1	0.251	0.376	0.137	0.539	0.139	0.526
2	0.148	0.517	0.170	0.479	0.085	0.672
3	0.096	0.639	0.086	0.669	0.092	0.651
4	0.047	0.820	0.016	1.000	0.117	0.584

Given that there are different influencing degrees on profits, the three products' respective weight is set as 0.4, 0.3, and 0.3. And then the obtained grey relational grade sequence between the mining objective and grey categories is $\tau=\{470, 0.552, 0.652, 0.803\}$. It is seen that the mining objective is in the fourth grey category, with its profit growth rate ranging from 0.30 to 0.40. The fact that the actual growth rate is 0.337, which falls within the obtained range, testifies the effectiveness of the mining model herein.

5. Conclusion

The paper discusses data mining of complex system and proposes a multi-index grey relational data mining model of complex system based on the grey system theory. By analyzing s string of grey category, index treatment, grey distance, grey relational coefficient and grey relational degree, the paper obtains the mining objective's area to which it belongs. The clear and easily-calculated model is testified to be effective and thus provides a new approach to address data mining of complex system.

References

[1] M. Li, Y. Tang and L. Zhou, "Data mining techniques and applications", China New Telecommunications, no. 22, **(2012)**, pp. 66-67+74.

[2] E. J. Khatib, R. Barco, A. G. Andrades, P. Muñoz and I. Serrano, "Data mining for fuzzy diagnosis systems in LTE networks", Expert Systems with Applications, vol. 42, no. 21, **(2015)**, pp. 7549-7559.

[3] Q. L. Hua, L. Sheng, C. X. Yang, C. Yan, Y. Xia and W. X. Yun, "Overview of Classification, Tools and Models of Data Mining", Modern Computer, no. 4, **(2013)**, pp. 17-21.

[4] L. Qifei and H. Hongtao, “The Applications Study of Data Mining in E-commerce”, Computer CD Software and Applications, no. 3, **(2012)**, pp. 195-196.
[5] S. Zhongxiang, P. Xiangjun, Y. Yuping and H. Yi. “Summarization on the Data Mining Application in Education Teaching”, Intelligent Computer and Applications, no. 2, **(2012)**, pp. 78-81.
[6] Z. Zhirong, H. Jie, H. Xiaodong and S. Xianming, “A Preliminary Survey on Data Mining in Application of Software Engineering”, Electronic Test, no. 9, **(2013)**, pp. 246-247+271.
[7] S. X. Ying and L. X. Jing, “Application of Data Mining in Stock Price Prediction”, Computer Simulation, vol. 29, no. 7, **(2012)**, pp. 375-378.
[8] X. P. Xiao, J. H. Wen and M. Xie, “Grey relational analysis and forecast of demand for scrap steel”, The Journal of Grey System, vol. 22, no. 1, **(2010)**, pp. 73-80.
[9] L. Sifeng, F. Zhigeng and Y. Y. Jie, “General grey numbers and its operations”, Grey Systems: Theory and Application, vol. 2, no. 3, **(2012)**, pp. 4-15.
[10] S. K. Nayak, J. K. Patro, S. Dewangan and S. Gangopadhyay, “Multi-objective Optimization of Machining Parameters During Dry Turning of AISI 304 Austenitic Stainless Steel Using Grey Relational Analysis”, Procedia Materials Science, vol. 6, **(2014)**, pp. 701-708.
[11] R. Vinayagamoorthy and M. A. Xavior, “Parametric Optimization on Multi-Objective Precision Turning Using Grey Relational Analysis”, Procedia Engineering, vol. 97, **(2014)**, pp. 299-307.
[12] J. B. Saedon, N. Jaafar, M. A. Yahaya, N. H. Saad and M. S. Kasim, “Multi-objective Optimization of Titanium Alloy through Orthogonal Array and Grey Relational Analysis in WEDM”, Procedia Technology, vol. 15, **(2014)**, pp. 833-841.

Authors

Xiaohui Wang, a Lecturer at Pingdingshan University, Pingdingshan, China. She received the B.Sc., M.Sc., degrees from Henan University and Xidian University, Kaifeng and Xian, China, in 2005 and 2009 respectively. Her research interests include: Intelligent Algorithm, and Networks Security.

Jianfen Liu, She received her B.Sc. from Henan Normal University (1995) and M.Sc. from Wuhan University of Science and Technology (2003). Now she is an associate professor at Pingdingshan University, Pingdingshan, China. Her current research interests include: Network Information Processing, and Intelligent Algorithm.