

Performance of computer examination items selection based on grey relational analysis

Yung-Fa Huang¹, Hsing-Chung Chen^{2*}, Pi-Ling Yen³

¹ Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan

² Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan

³ Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung, Taiwan

ABSTRACT

In this paper, we propose an improved Grey relational analysis (GRA) model to enhance the strategy of tests items selection. In general, the strategy of tests items selection is based on the item response theory (IRT) to select test items by the measuring information of target test items. Five ability levels are used to set out the target information for items selection. Simulation results show that the proposed improved GRA model with scaling factor by 1.5, under the number of items by 10, 15, 20, 25, and 30, performs better than both the average target information (ATI) and random methods in mean square error (MSE). Furthermore, the improved GRA model while the number of items is 20 could largely achieve lower MSE of test information error (TIE) by 0.8 comparing to the 2.6 and 4.2 of ATI and random methods, respectively.

Keywords: Grey system theory, Grey relational analysis, Item response theory, Target information, Item selection strategy.

OPEN ACCESS


Received: January 8, 2021

Revised: March 8, 2021

Accepted: April 7, 2021

Corresponding Author:

Hsing-Chung Chen
cdma2000@asia.edu.tw

 **Copyright:** The Author(s). This is an open access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted distribution provided the original author and source are cited.

Publisher:

[Chaoyang University of Technology](https://www.chaoyang.edu.tw/)

ISSN: 1727-2394 (Print)

ISSN: 1727-7841 (Online)

1. INTRODUCTION

Regardless of the teaching process of any learning institution, it is very important to develop a test item in order to confirm the learning effectiveness of students, that it could be used at all levels to evaluate learning effectiveness based on appropriate item selection and quantitative analysis (Crocker and Algina, 1986). In addition to the test of academic ability at the time of entrance to school to evaluate the effectiveness of their basic subject abilities, the appropriate test items for basic courses in elementary and middle schools is required adopting to evaluate students' learning effectiveness in each semester or quarter. Therefore, the design of test items is one of the important issues to achieve correct evaluation for schools (Fives and Barnes, 2016; Yu, 1993).

The item response theory (IRT) is widely used in many fields recently because it provides us the abilities of examinees and the difficulties of items simultaneously. Online IRT test systems are often equipped with the adaptive testing (Hirose and Aizawa, 2014). Such a testing method selects the most appropriate items to examinees automatically. However, the difficulty of items cannot adapt to the different group of the examinees. In order to enable students to quickly and effectively find the appropriate test items from the item bank during the subject test, the automatic selection of items is the main trend. Moreover, the computer item selection can achieve the learning evaluation in both of quantity and quality.

When both quantity and quality of the test items become larger, the network technology have been applied to personal computerized fitness tests (Fives and Barnes, 2016; Yu, 1993). However, in various examinations, group-based and paper-based tests are still widely used in Taiwan. In the multi-entry academic ability test or the middle

school examination, the test items are selected from a test-type item bank (He et al., 1998). However, how to select the test items to meet the special purpose of the examinees is an important research topic.

Grey system theory was proposed by China scholar Julong Deng in 1982 (Deng, 1992; Deng, 1996). Since its development and application, there have been many good research results (Yuan et al., 2017; Muddineni et al., 2020). Grey relational analysis (GRA) is mainly through the correlation between parameters, to find the required information from some known and ambiguous conditions, and then to understand the interactive relationship between the parameters. The target data of GRA is mainly based on less information. That is with less data and uncertainty to analyse quantification and serialization relationships from a multi-dimensional perspective. GRA is an overall comparison with a reference system (Younas et al., 2019). However, the comparison between two systems is a comparison without a reference system. The disadvantage of the comparison without a reference system is that it is easy to produce illusions and misunderstandings due to there is no reference object and the comparison environment is ignored. Therefore, this paper applied the GRA model as the parameter selection strategy for the test items. Moreover, the improved GRA model is proposed in order to effectively find the appropriate test items.

The traditional items selection methods such as Standard, Random, Middle Difficulty, Up and Down, and The Maximum Information Method, could quickly select the test items, but the selecting items cannot achieve the target information values (Yu, 1993). The traditional method of selecting items is arbitrary selection method (Yu, 1993) with using random number table, and randomly selecting 30 items from the item bank. Recently, the computer-based test items selection methods are popular used to reduce the test information error (TIE), which are defined by the difference between the selected test information function and the target information function.

The greedy algorithm was used as the item selection strategy by Sun et al. (1999). For each time, an item is added, if the tuning of the item information is calculated as 'negative', it means that the test information function is gradually approaching the target information function. The greedy item selection could reduce TIE compared with the traditional item selection method. However, its computational complexity is too high. Therefore, this paper proposed the improved GRA model which is the item selection strategy in order to effectively reduce TIE and reduce the complexity.

The main contributions of this paper are summarized as the following:

- In this paper, an improved adaptive GRA model is designed for effective relational analysis for multi-dimensional systems. The proposed improved GRA model consists of scaling factor to improve the adaptation of grey correlation coefficients.

- Three item selection strategies: Average Target Information (ATI), Random Selection, Improved GRA Selection are proposed to perform the optimization for items selections. The compilation processes for all item selection strategies are built to perform the optimization for computer examination items selections.
- Five ability levels are set out the target information for items selection. The proposed improved GRA can effectively select the test items with lowest MSE of TIE than ATI and Random methods for five ability levels.

The remainder of the paper is organized as follows. In Section 2, the GRA are depict for adaptive relational analysis for multi-dimensional sequences. Moreover, the proposed improved GRA models are described in this Section. Section 3 presents the details of the IRT. Section 4 describes the implementation of the item selection strategies. Section 5 shows the simulation results and a discussion of the results. Section 6 provides a conclusion.

2. GREY RELATIONAL ANALYSIS

The GRA is a measure of the degree of correlation between discrete sequences in grey system theory (Deng, 1996). It is different from the methods used with previous factors analysis. However, the GRA needs less data to perform the multi-factor analysis. With some analogous sequences, GRA performs the grey correlation generation (Deng, 1996).

There are $M+1$ sequences in the GRA space as

$$x_i(k) = (x_i(1), x_i(2), \dots, x_i(k), \dots, x_i(K)), \quad (1)$$

where $i = 0, \dots, M, k = 1, \dots, K$. That is,

$$x_0(k) = (x_0(1), x_0(2), \dots, x_0(k), \dots, x_0(K))$$

$$x_1(k) = (x_1(1), x_1(2), \dots, x_1(k), \dots, x_1(K))$$

...

$$x_M(k) = (x_M(1), x_M(2), \dots, x_M(k), \dots, x_M(K)). \quad (2)$$

Grey correlation coefficients can be obtained by

$$\gamma_{0i} = \frac{\Delta_{\min} + \zeta \cdot \Delta_{\max}}{\Delta_{0i}(k) + \zeta \cdot \Delta_{\max}} \quad (3)$$

where $i = 1, \dots, M, k = 1, \dots, K, j \neq i$. $x_0(k)$ is the reference sequence, $x_i(k)$ is a specific compared sequence.

$\Delta_{0i}(k) = |x_0(k) - x_i(k)|$ is the absolute value of the difference between the i th sequence and the reference sequence.

$\Delta_{\min} = \min_{\forall j \in i} |x_0(k) - x_j(k)|$ is the minimal value among all the subtracted sequences.

$\Delta_{\max} = \max_{\forall j \in i} |x_0(k) - x_j(k)|$ is the maximal value among all the subtracted sequences.

ζ is identification coefficient $\zeta \in (0, 1]$. Its value could be adjusted according to the practical situations. The magnitude of the identification coefficient ζ could be known from the actual mathematical verification that it will only change the magnitude of the relative value without affecting the ranking of its grey

correlations (Wen and Wu, 1996). Therefore, this paper set the value of ζ to 0.5.

When the grey correlation coefficient is obtained, if the number of grey correlation coefficients is too large, or the information is too scattered, it is not easy to compare (Wong and Lai, 2000). Therefore, the sum or average value of the correlation coefficients is called 'grey correlation degree', which is expressed as

$$\Gamma_i = \sum_{k=1}^K \gamma_{0i}(k) \quad (4)$$

where the improved grey correlation coefficients can be obtained by

$$\gamma_{0i}(k) = \frac{\Delta_{\min} + \zeta \cdot \Delta_{\max}}{[\Delta_{0i}(k)]^\lambda + \zeta \cdot \Delta_{\max}} \quad (5)$$

where λ is a scaling factor, $\lambda \in [1, 2]$.

3. ITEM RESPONSE THEORY

The classical test theory and IRT are the commonly used test theory (Yu, 1993; Hirose and Aizawa, 2014). The classical test theory is the earliest test theory among all test theories to establish empirical relationships between data. It mainly on estimating the reliability of the real scores of a test. The recent developed IRT aims on the lacking of classical test theory (Hirose and Aizawa, 2014). There are the following amendments and features in IRT:

1. The estimation of the parameters of contemporary test theory is not affected by the tester's sample.
2. The measuring error indexes are given for testers of different ability levels. Therefore, the ability value of the testers could be accurately estimated.
3. In addition to providing more accurate estimates for testers with different abilities, different ability estimates are also provided for testers with the same original score.
4. IRT is capable of measuring the subject's ability value for tests of homogeneity without being affected by the tests.
5. IRT uses item information and test paper information to evaluate the overall accuracy of the test paper and the test items.

IRT aims to achieve the effective testing by the tester in which the most important means is item information. The item information function proposed in the IRT depicts the information generated by the test under different ability levels. This information is used to select test items and to compare the relative performance of the tests. It is used as the main reference basis for establishing analysis and diagnostic tests. The definition of the test information function is expressed as

$$I_i(\theta) = \frac{(1.7a_i)^2(1-a_i)}{[c_i + e^{1.7a_i(\theta-b_i)}] + e^{-1.7a_i(\theta-b_i)}} \quad (6)$$

$I_i(\theta)$: the information value of the testee with ability level θ on the i th item.

a_i : Discrimination, range between 0 and 2.

b_i : Difficulty, usually only between ± 2 .

c_i : Guess.

From (6), it could be known that the closer the difficulty b_i is to the capability value θ , the larger the information value becomes. When the discrimination degree a_i is higher, the information value is larger. When the guess c_i is close to 0, the information value increases (Hirose and Aizawa, 2014).

4. TEST ITEMS SELECTION

In the item constructed based on the IRT, how to select an item from the item bank to form an item paper to meets the requirements of the item is an important issue. In this paper, we used the information function to select the items to reach the desired target information function of the test items. The steps of its compilation process were proposed by Lord (1980):

- Step 1: The target information function is assigned to information function required by the tester.
- Step 2: Select a test item from the item bank to make the sum of the information content of the selected test item to reach the target information function.
- Step 3: After each new item selection, the information content of the selected item need to be re-calculated.
- Step 4: Repeat the above steps until the test information function and the IFE of the information function reach the desired levels.

In terms of the information function, each item consists of a different amount of information $I_i(\theta)$ for different ability levels θ . When the information value is larger, it indicates that the item selection is appropriate. The information value in a test to different ability of the subject is accumulated for each item in the test. Therefore, with different test needs, different items could be selected from the item bank to achieve the goal of the test designer.

We set a target information values D_θ for each ability level θ . After the item selection is finished, the information values of the test formed by the selected item could be calculated as O_θ . Then, the Mean Square Error (MSE) of the TIE is defined as

$$MSE = \sum_{\theta=1}^K (D_\theta - O_\theta)^2 \quad (7)$$

where K is the number of ability levels.

4.1 Item Selection Strategy

Let the item bank \mathbf{T} be the set of test items t_i , $\mathbf{T} = \{t_1, t_2, \dots, t_M\}$, where M is the number of items in the item bank. Assume that we have a target paper \mathbf{S} , $\mathbf{S} \subseteq \mathbf{T}$, $\mathbf{S} = \{s_1, s_2, \dots, s_i, \dots, s_N\}$, where s_i is the i th selected item for the target paper and N is the number of items for the target paper. Then, the information value of test item t_i at each ability level θ could be expressed by

$$t_i(\theta) = \frac{(1.7a_i)^2(1-c_i)}{[c_i + e^{1.7a_i(\theta-b_i)}][1 + e^{-1.7a_i(\theta-b_i)}]^2} \quad (8)$$

where $1 \leq i \leq M$, $-r \leq \theta \leq r$, r is a positive number, $t_i \in \mathbf{T}$. For the target items paper, the information value of the i th selected item s_i can be obtained by

$$s_i(\theta) = \frac{(1.7a_i)^2(1-c_i)}{[c_i + e^{1.7a_i(\theta-b_i)}][1 + e^{-1.7a_i(\theta-b_i)}]^2} \quad (9)$$

where $1 \leq i \leq N$, $s_i \in \mathbf{S}$, $-r \leq \theta \leq r$, $\theta \in Z$. In this paper, we collected the item bank set \mathbf{T} , and set the target information value D_θ ($-r \leq \theta \leq r$) for each ability level with the total number of selected items by N .

4.2 Average Target Information (ATI)

This item selection method is based on item-by-item selection. That is, each item is selected through the following three steps. First, the average information value of the i th selected item is defined by

$$d_\theta^i = \frac{D_\theta - \sum_{j=1}^i s_j(\theta)}{N'} \quad (10)$$

where $1 \leq i \leq N$, $-r \leq \theta \leq r$. In the selection of the i th item, the procedure of each step is as follows:

Step 1: Find the ATI values after selecting the i th item, d_θ^i in (10). With the selection of the i th item, the ATI value can be obtained by

$$d_\theta^i = \frac{D_\theta}{N} \quad (11)$$

where $-r \leq \theta \leq r$, θ is the ability level, r is a positive number.

Step 2: Find the sum of the absolute value of the difference between each test item and the ATI value in each item in the item bank, by

$$\Delta E_j = \sum_{\theta=-r}^r |d_\theta^i - t_j(\theta)| \quad (12)$$

where $1 \leq j \leq M$. The item of the minimal value in (15) is selected by

$$l = \arg \min_j (\Delta E_j) \quad (13)$$

Step 3: Perform the selection action, and select the items selected in (13) into the target test paper, that is, $s_i = t_l$. (14)

The selected items are removed from the item bank. The steps 1 to 3 are repeated for N times. Then the selected item \mathbf{S} is output for the test.

4.3 Random Selection

This method of item selection is also based on item-by-item selection. That is, each item is selected through the following three steps. But this method does not take into account the amount of information in the test items. The

information value will approach the average information value of the item bank \mathbf{T} by that is,

$$O_\theta^R = \frac{\sum_{j=1}^M t_j(\theta)}{M} \quad (15)$$

In the i -th item selection, the following three steps are performed.

Step 1: Take the system time, and the random number seeds, to select a uniform random number to the l th item $1 \leq l \leq M$. Select the selected test items into the target test paper, by $s_i = t_l$ in (14).

Step 2: If the number of selected items is not reached, $i < N$, go to Step 1, otherwise stop.

Step 3: Output the target test paper \mathbf{S} .

4.4 Improved Grey Relational Item Selection

The target information values are set to D_θ by reference sequence $x_0(k) = (x_0(1), x_0(2), \dots, x_0(k), \dots, x_0(K))$. The information values of the item bank $\mathbf{T}(\theta) = \{t_1(\theta), t_2(\theta), \dots, t_M(\theta)\}$ are set as the information sequences by $X(k) = \{x_i(k)\}$, $i = 1, \dots, M$. In the improved GRA strategy, the steps for the i th item selection are as follows.

Step 1: Find the size of the grey correlation coefficient. From the improved grey correlation coefficient of (5), the grey correlation coefficient of the i -th item is

$$\gamma_i = \frac{\Delta_{\min} + \zeta \cdot \Delta_{\max}}{[\Delta_i(k)]^\lambda + \zeta \cdot \Delta_{\max}} \quad (16)$$

where $i = 1, \dots, M$, $k = 1, \dots, K$, $\zeta = 0.5$, $\Delta_i(k) = |x_0(k) - x_i(k)|$, $\Delta_{\min} = \min_{j \in I} |x_0(k) - x_j(k)|$,

$\Delta_{\max} = \max_{j \in I} |x_0(k) - x_j(k)|$, and $\lambda = 1.5$.

Step 2: Find the sum of the grey correlation coefficients of the information sequence x_i of the i th item t_i by

$$\Gamma_i = \sum_{k=1}^K \gamma_i(k), \quad 1 \leq i \leq M. \quad (17)$$

The item of the maximal grey correlation coefficients in (17) is selected by

$$l = \arg \max_i (\Gamma_i) \quad (18)$$

Step 3: Perform the selection action, and select the test items selected in (18) into the target test paper, that is,

$$s_i = t_l \quad (19)$$

Step 4: After the i -th grey-related item selection, if $i < N$, a new reference sequence is set. The new target information value is defined as the difference between the target value and the test paper information value of the selected item at each ability level by

$$D_\theta^i = D_\theta - s_i(\theta) \quad (20)$$

where s_i is the selected i -th item. Then the $(i + 1)$ th item is selected by step 1. If $i = N$, the selection procedure is stop and the target test paper S is obtained.

5. SIMULATION RESULTS

In this paper, an item bank is adopted for the computerized fitness test system for the natural sciences of the National Primary School in Taiwan (Sun et al., 1999). The total number of item banks is $M = 320$. In order to evaluate the effectiveness of the grey correlation problem selection strategy, we select $N = 10, 15, 20, 25, 30$ items from 320 different items, and set the target information values $D_\theta (D_{-2}, D_{-1}, D_0, D_1, D_2)$ at different ability levels (θ). The target information value $D_\theta (-2 \leq \theta \leq 2)$ is shown in Table 1.

Table 1. The target information value $D_\theta (-2 \leq \theta \leq 2)$ for different number of items selection

N	D_θ				
	Ability levels (θ)				
	-2	-1	0	1	2
10	1	2	3	3	2
15	1	3	4	4	2
20	2	4	6	5	3
25	2	4	7	6	3
30	2	5	8	7	4

The simulation of the information values of the selected items according to the computer item selection strategies is performed by Matlab programming language. The TIE of the GRA selection, ATI method and Random method are performed to compare the target information value as shown in Fig. 1.

Fig. 1 shows the comparison between the test items selected by the three methods and the target information values D_θ at different ability levels (θ) with different number of items $N = 10, 20, 30$. From Fig. 1, it is observed that the information values by the Random method is mostly the lowest, while the results of the GRA method are the closest ones to the target information values than the ATI method and the Random method. Therefore, the improved GRA performs the best item selection on different ability levels.

The MSE of (7) is further calculated, and the MSE comparison of the three methods is obtained as shown in Fig. 2 at $N = 10, 15, 20, 25, 30$ items. From Fig. 2, it can be seen that except for $N = 15$, the GRA method and ATI are equal performance for the MSE. However, the GRA method is better than the ATI method in other N values, and the GRA method and the ATI method should be better than the random selection method. However, when $N = 25$, the ATI method is worse than Random method, which should be caused by a special situation of this item bank.

6. CONCLUSION

In this paper, the computerized test items selection is performed based on the information of the test theory to achieve the requirements of the tester's test target information values. The improved GRA model is proposed as item selection strategy to enhance the IFE of the item selection. The number of items of 10, 15, 20, 25, and 30 items are selected from the 320 item bank, respectively. The target information values are set at five ability levels. Simulation results show that the MSE of the TIE of the improved GRA method outperforms the ATI method and the Random method with the lowest MSE of the item selection.

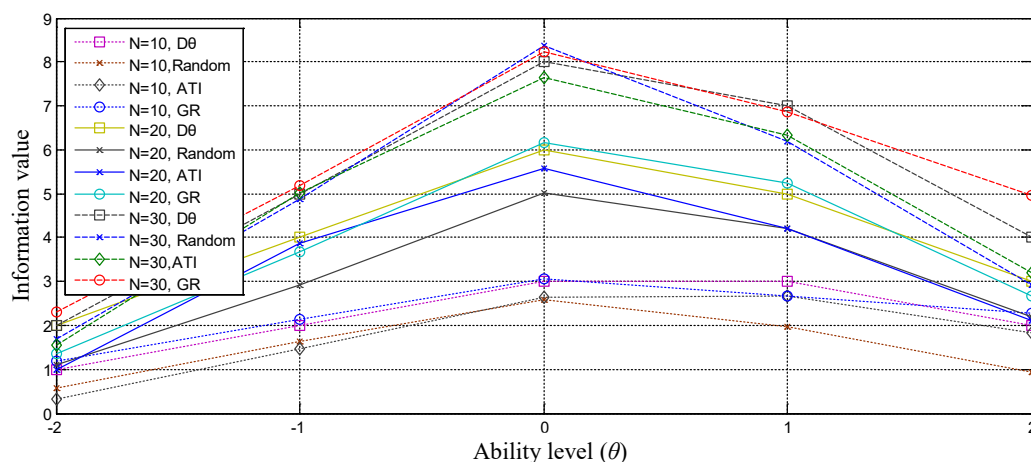


Fig. 1. Comparison of the information value of different item selection strategies at each ability level

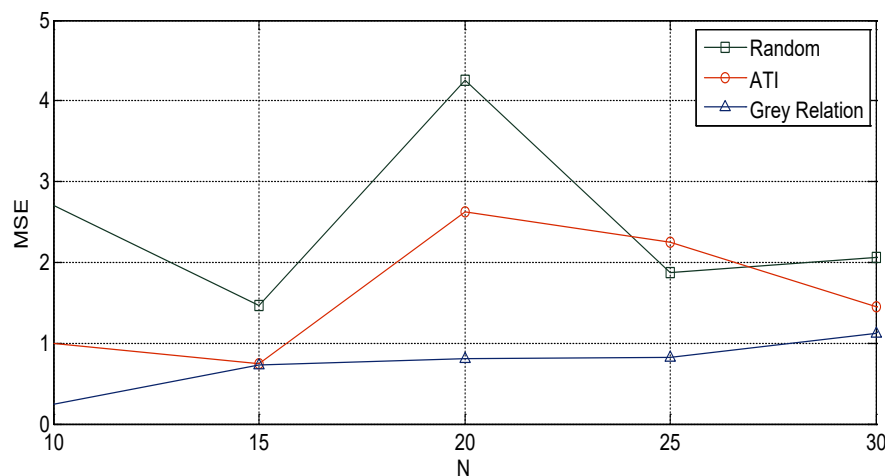


Fig. 2. Comparison of MSE information value error of different item selection strategies

ACKNOWLEDGMENT

This paper is partly supported by the Ministry of Science and Technology (MOST), Taiwan, with the grant number MOST 108-2221- E-324-010-.

REFERENCES

- Crocker, L., Algina, J. 1986. Introduction to classical and modern test theory. Holt, Rinehart and Winston. Florida. U.S.A.
- Deng, J. 1992. Grey system theory tutorial. Huazhong University of Science and Technology Press. Wuhan. China.
- Deng, J. 1996. Basic method of grey system. Huazhong University of Science and Technology Press. Wuhan. China.
- Fives, H., Barnes, N. 2016. Informed and uninformed naïve assessment constructors' strategies for item selection, Journal of Teacher Education. <https://doi.org/10.1177/0022487116668019>.
- He, R., Du, L., Zhuang, Q. 1998. Improved grey prediction computerized suitability test selection items. Proceeding of Seventh International Computer-aided Teaching Symposium, 393–400.
- Hirose, H., Aizawa, Y. 2014. Automatically growing dually adaptive online IRT testing, Proceeding of International Conference of Teaching, Assessment and Learning (TALE), 528–533.
- Lord, F.M. 1980. Applications of item response theory to practical testing problems. Routledge. Taylor & Francis. New York. U.S.
- Muddineni, V.P., Bonala, A.K., Sandepudi, S.R. 2020. Grey relational analysis based objective function optimization for predictive torque control of induction machine. Proceeding of 2020 IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020).
- Sun, G., Chen, Y., Lai, Y., Xie, K., Chen, X. 1999. Using greedy algorithms as a profitable item selection strategy. Proceeding of National Computer Conference. Tamkang University. Taipei, Taiwan, 379–386.
- Wen, K.-L., Wu, J.H. 1996. On distinguishing coefficient & relational grade. The Journal of Grey System, 8, 11–18.
- Wong, C.C., Lai, H.R. 2000. A new grey relational measurement. The Journal of Grey System, 12, 341–346.
- Younas, M., Jaffery, S.H.I., Khan, M., Khan, M.A., Ahmad, R., Mubashar, A., Ali, L. 2019. Multi-objective optimization for sustainable turning Ti6Al4V alloy using grey relational analysis (GRA) based on analytic hierarchy process (AHP). The International Journal of Advanced Manufacturing Technology, 105, 1175–1188.
- Yu, M. 1993. Computerized adaptability test. Study Information, 10, 5–9.
- Yuan, C., Zhang, Y., Xu, N., Xu, J. 2017. Grey relational analysis on the relation between China's GDP and defense expenditure, Proceeding of 2017 IEEE International Conference on Grey Systems and Intelligent Services (GSIS).