Research on Physical Fitness Test Data Mining and Analysis Based on Apriori Algorithm

Chengxiang Shi¹*, Lei Liu¹

¹ Department of Mathematics and Information Engineering, Chongqing University of Education, Chongqing, 400060, China

*Corresponding author’s e-mail: 6765443@qq.com

Abstract. Based on the analysis of the characteristics of Apriori algorithm, this paper uses the iterative method of layer by layer search to find out the relationship of item sets in the students' body measurement database to form the mining model of association rules. The model maximizes the generation of candidate sets, reduces the size of frequent sets, and achieves good performance. At the same time, the average GM (1,1) model is constructed by the accumulation of the original sequence and the least square estimation of the parameters, as well as the whitening differential equation form, to predict the trend of College Students' physique change in the future, and the research shows that the error of the model is small and the precision is high.

1. Introduction

College students are the main force of social development. In order to solve the problem of College Students’ physique, through the in-depth study and analysis of the data mining results, we can understand the correlation between College Students’ physique and health test items more comprehensively, it is of great practical significance to improve the physical health of college students, to promote physical education in Colleges, to improve the quality of teaching and the construction of indoor and outdoor venues.

2. Index system construction

Based on the 2015-2018 physical test data of Chongqing University of Education, 12255 samples were taken to construct the physical test index system. Eliminate the invalid data and make descriptive statistical analysis on the eliminated data. As shown in table 1, table 2.

| Table 1. 2015-2018 Average value of all body measurement indexes of female students |
|------------------|----------------|----------|---------|------------------|------------------|------------------|
| Grade | Height (cm) | Weight (kg) | Vital capacity (ml) | 50 meter (s) | Standing long jump (cm) | Sit-and-reach (cm) | 800 meter (s) | Sit-up (pcs) |
|-------|-------------|-------------|---------------------|-------------|------------------------|-------------------|-------------|-------------|
| 2015  | 159.114     | 51.548      | 2407.65             | 10.074      | 166.570                | 19.001            | 242.730     | 31.460      |
| 2016  | 160.200     | 51.390      | 2540.00             | 9.876       | 164.600                | 16.629            | 248.820     | 34.190      |
| 2017  | 159.252     | 52.429      | 2422.53             | 9.893       | 161.940                | 16.681            | 242.420     | 34.780      |
| 2018  | 160.455     | 51.720      | 2574.52             | 9.585       | 161.660                | 16.540            | 245.540     | 36.070      |
Table 2. 2015-2018 Average value of male students’ body measurement index data

| Grade | Height (cm)  | Weight (kg)  | Vital capacity (ml) | 50 meter (s) | Standing long jump (cm) | Sit-and-reach (cm) | 1000 meter (s) | Pull-up (pcs) |
|-------|--------------|--------------|---------------------|-------------|------------------------|-------------------|---------------|-------------|
| 2015  | 171.930      | 61.968       | 3707.257            | 7.983       | 220.729                | 15.022            | 252.004       | 8.540       |
| 2016  | 171.419      | 62.290       | 3631.209            | 8.025       | 220.818                | 13.867            | 250.597       | 7.273       |
| 2017  | 173.025      | 62.704       | 3789.969            | 8.405       | 217.670                | 14.063            | 257.771       | 9.635       |
| 2018  | 172.688      | 61.919       | 3862.011            | 7.714       | 216.607                | 14.895            | 255.000       | 10.421      |

From table 1, it can be seen that the average change trend of each index of female students is: the 50m run and the sit-and-reach gradually become smaller; the sit-up gradually become larger; the height, weight, vital capacity, standing long jump and 800m run oscillate. From table 2, we can get the average change trend of various indicators of boys: standing long jump gradually becomes smaller, pull-up, vital capacity generally gradually becomes larger, height, weight, 50m run, sit-and-reach, 1000m run oscillation changes, the trend is uncertain.

3. Analysis of body measurement data based on Apriori algorithm of association rules

3.1. Apriori algorithm of association rules data analysis

Apriori algorithm uses candidate item set to find frequent item set, which is the most influential algorithm to mine frequent item set of Boolean association rules. In the case of female students, the minimum rule confidence is 80%, and rules with gain less than 1 are eliminated. Finally, the results are visualized as a mesh for overall analysis, as shown in table 3 and figure 1.

Table 3. Effective rule (girls)

| Consequent | The aforesaid | Rule identification | Examples | Support | Confidence | Rule support | Gain |
|------------|---------------|---------------------|----------|---------|------------|--------------|------|
| pass in 50m| the level of 800m is good | 3 | 1,031 | 11.395% | 91.950% | 10.477% | 1.233 |
| pass in 50m| the level of standing long jump is good | 1 | 913 | 10.091% | 90.909% | 9.173% | 1.219 |
| pass in 800m| failed in 50m | 9 | 2,213 | 24.458% | 81.654% | 19.971% | 1.060 |
| pass in 800m| failed in vital capacity | 5 | 1,127 | 12.456% | 80.745% | 10.057% | 1.048 |
| pass in sit-up| pass in standing long jump | 13 | 6,414 | 70.889% | 85.859% | 60.864% | 1.013 |
| pass in sit-up| pass in 50m | 14 | 6,750 | 74.602% | 85.807% | 64.014% | 1.012 |
| pass in sit-up| the level of sit-and-reach is good | 8 | 1,589 | 17.562% | 85.651% | 15.042% | 1.010 |
| pass in sit-up| pass in sit-and-reach | 12 | 4,748 | 52.476% | 85.320% | 44.772% | 1.006 |
| pass in sit-up| pass in vital capacity | 15 | 6,856 | 75.774% | 85.298% | 64.633% | 1.006 |
| pass in sit-up| pass in 800m | 16 | 6,968 | 77.011% | 85.146% | 65.573% | 1.004 |
The purpose of this paper is to study the promotion of each index in order to improve the physical quality of students. For example, rule mark 3, under the condition of good 800m run grade, the probability of passing 50m run grade is 91.95%, which shows that 800m run can promote 50m run. The thickness and depth of the middle line in the figure represent the strength of the connection. It can be seen intuitively that in addition to sitting forward bending, other indicators of girls have strong connection. Repeat the above operations to get the association between boys' indicators shown in table 4 and figure 2.

The aforesaid

| Consequent                  | The aforesaid                  | Rule identification | Examples | Support    | Confidence | Rule support | Gain   |
|-----------------------------|--------------------------------|---------------------|----------|------------|------------|--------------|--------|
| pass in 50m                  | failed in pull-up              | 8                   | 1,754    | 54.693%    | 95.838%    | 52.417%      | 1.028  |
| pass in 50m                  | pass in 1000m                  | 10                  | 2,171    | 67.696%    | 95.762%    | 64.827%      | 1.027  |
| pass in 50m                  | failed in standing long jump   | 5                   | 677      | 21.110%    | 95.421%    | 20.143%      | 1.024  |
| pass in 50m                  | failed in vital capacity       | 2                   | 456      | 14.219%    | 95.395%    | 13.564%      | 1.024  |
| pass in 50m                  | pass in standing long jump     | 12                  | 2,252    | 70.221%    | 94.583%    | 66.417%      | 1.015  |
| pass in 50m                  | pass in sit-and-reach          | 9                   | 1,786    | 55.691%    | 94.513%    | 52.635%      | 1.014  |
| pass in 50m                  | pass in vital capacity         | 11                  | 2,216    | 69.099%    | 93.953%    | 64.920%      | 1.008  |
| pass in 50m                  | pass in pull-up                | 7                   | 1,003    | 31.275%    | 93.619%    | 29.280%      | 1.004  |
| pass in 50m                  | the level of sit-and-reach is good | 3                 | 478      | 14.905%    | 93.515%    | 13.938%      | 1.003  |
| pass in 50m                  | failed in 1000m                | 4                   | 620      | 19.333%    | 93.387%    | 18.054%      | 1.002  |
It can be seen intuitively that boys' standing long jump, 50m run and 1000m run have a strong connection.

4. Metrological analysis of body measurement data based on grey prediction model[4]

4.1. Grey prediction model

Based on the distinct characteristics of small sample, poor data and various body measurement indexes of male and female students, it is more suitable to use the grey prediction EGM model. Take 50 meters for girls as an example:

- Make original sequence,
  \[ X^{(1)} = \{ x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), x^{(1)}(4) \} = \{ 10.074, 19.95, 29.843, 39.428 \} \]

- Calculate 1-AGO,
  \[ X^{(0)} = \{ x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4) \} = \{ 10.074, 9.876, 9.893, 9.585 \} \]

- Calculate nearest mean generation, \( z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k - 1) \) \hspace{1cm} (1)

- Calculate \( Z^{(1)} = \{ z^{(1)}(2), z^{(1)}(3), z^{(1)}(4) \} = \{ 15.012, 24.8965, 34.6855 \} \)

- Calculate \( B = \begin{bmatrix} -z^{(1)}(2) & 1 & -z^{(1)}(3) & -z^{(1)}(4) \end{bmatrix} \), \( Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ x^{(0)}(4) \end{bmatrix} = \begin{bmatrix} 9.876 \\ 9.893 \\ 9.585 \end{bmatrix} \)

- Calculate \( \hat{a} = (B^TB)^{-1}B^TY = \begin{bmatrix} 0.0148 \\ 10.1521 \end{bmatrix} \) \hspace{1cm} (2)

- Definite model, \( \frac{dx^{(1)}}{dt} + 0.0148x^{(1)} = 10.1521 \) \hspace{1cm} (3)

- Calculate time response,
  \[ x^{(1)}(k+1) = (x^{(1)}(1) - \frac{b}{\hat{a}})e^{-\hat{a}k} + \frac{b}{\hat{a}} = (10.074 - \frac{10.1521}{0.0148})e^{0.0148k} + \frac{10.1521}{0.0148} \] \hspace{1cm} (4)

- Calculate \( \hat{X}^{(1)} = \{ \hat{x}^{(1)}(1), \hat{x}^{(1)}(2), \hat{x}^{(1)}(3), \hat{x}^{(1)}(4) \} = \{ 10.074, 20.004, 29.787, 39.428 \} \)

- Restore the calculated analog value, \( \hat{x}^{(0)}(k) = a^{(1)}x^{(1)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k - 1) \) \hspace{1cm} (5)

- Calculate \( \hat{X}^{(0)} = \{ \hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \hat{x}^{(0)}(3), \hat{x}^{(0)}(4) \} = \{ 10.074, 9.930, 9.784, 9.640 \} \)

- Check the error, \( e^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k) = \{ 0, -0.054, 0.109, -0.055 \} \)
• Calculate mean relative error, \[ \Delta = \frac{1}{3} \sum_{k=2}^{4} \Delta_k = \frac{1}{3} \sum_{k=2}^{4} \left| \epsilon(k) \right| = 0.741\% \] (7)

4.2. Prediction analysis

Repeat the above steps to get the predicted values of other indicators and the corresponding mean relative errors, as shown in table 5 and table 6.

Table 5. 2019-2021 prediction of the average value of each physical measurement index (girls)

| Grade | Height (cm) | Weight (kg) | Vital capacity (ml) | 50 meter (s) | Standing long jump (cm) | Sit-and-reach (cm) | 800 meter (s) | Sit-up (pcs) |
|-------|-------------|-------------|---------------------|-------------|------------------------|-------------------|-------------|-------------|
| 2019  | 160.225     | 52.175      | 2547.689            | 9.500       | 159.807                | 16.528            | 242.310     | 36.940      |
| 2020  | 160.353     | 52.341      | 2565.571            | 9.359       | 158.367                | 16.484            | 240.686     | 37.950      |
| 2021  | 160.481     | 52.507      | 2583.579            | 9.222       | 156.939                | 16.440            | 239.074     | 38.984      |

Table 6. 2019-2021 prediction average relative error of the average value of each body measurement index (girls)

| Height (cm) | Weight (kg) | Vital capacity (ml) | 50 meter (s) | Standing long jump (cm) | Sit-and-reach (cm) | 800 meter (s) | Sit-up (pcs) |
|-------------|-------------|---------------------|-------------|------------------------|-------------------|-------------|-------------|
| MRe         | 0.300%      | 0.747%              | 2.406%      | 0.741%                 | 0.324%            | 0.258%      | 0.863%      | 0.431%      |

Height, weight, vital capacity and sit-ups showed a slow growth trend; standing long jump and sitting forward bending gradually decreased; 50m and 800m time gradually shortened. From table 6, it can be seen that the mean relative error of the average prediction is larger than 1% except for the mean relative error of vital capacity. The mean relative error of the remaining index prediction is smaller than 1%. It shows that the EGM model is an ideal model for predicting the development trend of physical fitness. By analogy, the predicted values and mean relative errors of boys' indicators is shown in table 7 and table 8:

Table 7. 2019-2021 prediction of the average value of each physical measurement index (boys)

| Grade | Height (cm) | Weight (kg) | Vital capacity (ml) | 50 meter (s) | Standing long jump (cm) | Sit-and-reach (cm) | 1000 meter (s) | Pull-up (pcs) |
|-------|-------------|-------------|---------------------|-------------|------------------------|-------------------|--------------|--------------|
| 2019  | 173.648     | 63.269      | 3996.575            | 7.748       | 214.179                | 15.340            | 258.860      | 12.589       |
| 2020  | 174.287     | 63.587      | 4120.608            | 7.603       | 212.121                | 15.907            | 261.095      | 14.885       |
| 2021  | 174.928     | 63.907      | 4248.491            | 7.461       | 210.082                | 16.495            | 263.349      | 17.599       |

Table 8. 2019-2021 prediction mean relative error of the average value of each physical measurement index (boys)

| Height (cm) | Weight (kg) | Vital capacity (ml) | 50 meter (s) | Standing long jump (cm) | Sit-and-reach (cm) | 1000 meter (s) | Pull-up (pcs) |
|-------------|-------------|---------------------|-------------|------------------------|-------------------|--------------|--------------|
| MRe         | 0.251%      | 0.427%              | 0.530%      | 2.936%                 | 0.149%            | 0.970%       | 0.867%       | 4.478%       |

Height, weight, vital capacity and sit-and-reach show a slow growth trend; the pull-up increase is large; the standing long jump is gradually reduced; the time of 50m is gradually shortened, but the
time of 1000m is gradually increased. From table 8, except the average relative error of 50m, the average relative error of the remaining index prediction is less than 1%, which also shows that the EGM model is an ideal model to predict the development trend of physical fitness.

5. Summary
Based on association rules and mean GM (1,1) model, the problems of association degree and average value prediction of physical fitness test indexes are effectively solved, and the effective rules and prediction equations are given, and the model has the characteristics of small error and high accuracy. Through the research of this paper, we can make the school and students understand their physique and health more clearly. At the same time, according to the test results of each student, we can choose the exercise strategy pertinently, make the exercise of the students more scientific, so as to achieve the purpose of improving the students' physique and health level in an all-round way. In this study, the accuracy of the grey prediction model is high, but the association degree between the various body measurement indexes using the Apriori algorithm of association rules needs to be improved.

Acknowledgment
This research was partly financially supported through grants from the Scientific and Technological Research Program of Chongqing Municipal Education Commission (No. KJQN201801610), the Scientific Research Project of Chongqing University of Education (No. KY201701A), the Chongqing Urban and Rural Teacher Education Research Center, Chongqing Municipal Humanities and Social Sciences Key Research Base (No.18JDZDWT03), the Education and Teaching Reform Research Project of Chongqing University of Education (No. JG201922).

References
[1] Zhao Guorui, Hao Shiyu, Tian Zhenming. Analysis of College Students' physical fitness based on correlation theory and grey model [J]. Journal of Jiamusi University (NATURAL SCIENCE EDITION), 36 (05): 141-143 + 172.
[2] Lin Yijun, Shen Liangzhong. (2018) Research on the performance analysis of College Students' physical test based on decision tree technology [J]. Computer knowledge and technology.
[3] Shi Jianhua. (2018) Research on the correlation between BMI and physical fitness test index of college students [D]. Fujian Normal University.
[4] Zhao tjin, Zou Hua. (2018) A study on the relationship between the results of physical fitness test and physical education teaching of Dali University Students [J]. Neijiang science and technology, v.39; no.279 (02): 108-109.
[5] He Wei, Wu Yingman, Yuan Wei. A study on the relationship between the results of physical fitness test and physical education [J]. Journal of China Institute of labor relations (1): 125-128.
[6] Wu Guoyun, Fu Zhikai. (2018) Correlation analysis of BMI, body function index and quality index of College Students -- Taking Xinzhou Normal University as an example [J]. Stationery and technology.
[7] Yuan Yunhua. Analysis and Research on the results of physical fitness test of students in Wuxi Higher Vocational College [D].
[8] Xu tingwen. (2019) Analysis of physical health test data of college students and Research on Countermeasures -- Taking the sample data of 2018 as an example [J]. Stationery and technology.
[9] Wu Jinwei. A study on the correlation between physical test scores and health of college students [J]. Journal of Jilin Institute of physical education, V.34; No.150 (02): 44-48.