Attribute Coordinate Comprehensive Evaluation Model Combining Principal Component Analysis

Xiaolin Xu[1(✉)], Yan Liu[1(✉)], and Jiali Feng[2]

[1] Shanghai Polytechnic University, Shanghai, China
`xlxu2001@163.com, liuyan@sspu.edu.cn`
[2] Shanghai Maritime University, Shanghai, China
`jlfeng@shmtu.edu.cn`

Abstract. Attribute coordinate comprehensive evaluation method features subjective weighting in which the weights of indicators are determined by evaluators, which possibly leads to the arbitrariness in setting the weights. When there are many indicators, it is difficult to accurately judge if the sample is better or worse than others. To address the problem, this paper applies principal component analysis on the attribute coordinate comprehensive evaluation method. When there are many indicators, they can be reduced to new indicators with related meanings given through the method of principal component analysis. With the simplification, it will greatly facilitate experts to rate samples, which is the paramount basis that provides the preference of experts for the attribute coordinate comprehensive evaluation method to further calculate all the satisfaction degrees of objects to be evaluated. Experimental results show the advantages of the improved algorithm over the original algorithm.

Keywords: Attribute coordinate comprehensive evaluation
Principal component analysis · Barycentric coordinates
Global satisfaction degree

1 Introduction

In regards to comprehensive evaluation, the most important problem needed to address is how to set the weight of each evaluated indicator. The setting of weights can fall into two categories. One is the subjective weighting, such as AHP [1, 2], and the other is the objective weighting, such as the least square method and the principal component analysis [3, 4]. The two types have their own advantages and disadvantages. Subjective weighting is that the weights are given by experts and could be arbitrary in some cases, while objective weighting is not able to reflect the experiences or preferences of experts. Attribute coordinate comprehensive evaluation, belonging to the former, whose characteristic is that it can construct the corresponding psychological preference curve through evaluators rating the sample data in light of their own experiences or

preferences, has made certain progress both in theory and practice [5–13]. However, when indicators are many, it is difficult for experts to accurately distinguish satisfactory samples from unsatisfactory samples, which might result in arbitrary ratings on some samples. To address the obstacle, the principal component analysis method is used to reduce the number of indicators and give the related meanings of new indicators, so it is easier for experts to rate on samples with new indicators.

This paper first introduces the steps of simplification of indicators by means of the principal component analysis, then explores the core idea of the attribute coordinate comprehensive evaluation method, and next elaborates the process of combining the two methods through the simulation and the comparison of results before and after the model is improved.

2 Reduction of Indicators by Principal Component Analysis

Principal component analysis is a method of dimensionality reduction in mathematics. The basic idea is to try to make the original indicators $X_1, X_2, \ldots X_t$ (for example, there are t indicators) recombined into a set of relatively unrelated comprehensive indicators F_m with fewer numbers than the number of original indicators. The specific steps of the principal component analysis are as follows:

(1) **Calculate the covariance matrix**

Calculate the covariance matrix $s = (s_{ij})_{p \times p}$ of sample data

$$s_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad i,j = 1, 2, \ldots, p \tag{1}$$

Among them, s_{ij} $(i, j = 1, 2, \ldots, p)$ is the correlation coefficient between the original variable x_i and x_j. p is the number of indicators. n is the number of samples. $\bar{x}_i$ and $\bar{x}_j$ is respectively the mean of values of indicator i and j. x_{ki} is the value of indicator i of a certain sample, and x_{kj} is the value of indicator j of a certain sample.

(2) **Calculate the eigenvalues λ_i of S and orthogonal unit eigenvectors a_i.**

The first m larger eigenvalues of S, $\lambda 1 \geq \lambda 2 \geq \ldots \lambda m > 0$, is the variance of the first m principal components, and the unit eigenvector a_i corresponding to λ_i is the coefficient of the principal component F_i, and then the ith principal component Fi is:

$$F_i = a_i X \tag{2}$$

The variance (information) contribution rate of principal components reflects the information magnitude, γ_i is:

$$\gamma_i = \lambda_i / \sum_{i=1}^{m} \lambda_i \tag{3}$$

(3) **Determine the principal components**

The final principal components to be selected are $F1$, $F2$, … Fm, and m is determined by the cumulative contribution rate of variance $G(m)$.

$$G(m) = \sum_{i=1}^{m} \lambda_i / \sum_{k=1}^{p} \lambda_k \tag{4}$$

When the cumulative contribution rate is greater than 85%, it will be considered enough to reflect the information of the original variables, and m is the extracted first m principal components.

(4) **Calculate the load of the principal components**

The principal component load reflects the degree of correlation between the principal component F_i and the original variable X_j, and the load $l_{ij}(i = 1, 2, …, m; j = 1,2, …, p)$ of the original variable X_j ($j = 1,2, … p$) on the principal component F_i ($i = 1, 2, …, m$) is:

$$l(F_i, X_j) = \sqrt{\lambda_i} a_{ij} (i = 1, 2, …, m; j = 1, 2, …, p) \tag{5}$$

(5) **Calculate the scores of the principal components**

The scores on the m principal components of the sample:

$$F_i = a_{1i}X_1 + a_{2i}X_2 + \cdots + a_{pi}X_p \qquad i = 1, 2, …, m \tag{6}$$

(6) **Select the principal components and give the new meanings**

Provide the new meaning of the new evaluation indicator F_i ($i = 1, 2, …, m$) for experts to rate on the new samples.

3   Attribute Coordinate Comprehensive Evaluation Model

3.1   Explore Barycentric Coordinates Reflecting Evaluators' Preference Weight

Attribute coordinate comprehensive evaluation method combines machine learning with experts' ratings on sample data. Set T_0 to be the critical total score, T_{max} the largest total score, we evenly select several total scores: T_1, T_2, … T_{n-1} from (T_0, T_{max}) regarding the curve fitting requirements, and then select some samples on each total score $T_i(i = 1, 2, 3 … n - 1)$ and rate them according to experts' preference or experiences, which is taken as the process of the learning of samples, so as to get the barycentric coordinate for T_i ($i = 1, 2, 3 … n - 1$) according to (7).

$$b\left(\{f^h(z)\}\right) = \left(\frac{\sum\limits_{h=1}^{t} v_1^h f_1^h}{\sum\limits_{h=1}^{t} v_1^h}, \ldots, \frac{\sum\limits_{h=1}^{t} v_m^h f_m^h}{\sum\limits_{h=1}^{t} v_m^h}\right) \tag{7}$$

Where, $\{f_k, k = 1, \ldots s\} \subseteq S_T \cap F$ is the set for sample f_i with the total score equal to T. In Formula (7), $b(\{v^h(z)\})$ is the barycentric coordinate of $\{v^h(z)\}$, $\{f^h, h = 1, \ldots t\}$ is the values of indicators of t sets of samples the evaluator Z selects from $\{f_k\}$, $\{v^h(z)\}$ is the ratings (or taken as weight) the evaluator gives on the samples.

3.2   Calculate the Most Satisfactory Solution

Use the interpolation formula $G_j(T) = a_{0j} + a_{1j} T + a_{2j} T_2 + \ldots + a_{n+1j} T_{n+1}$ and barycentric coordinates obtained above to do curve fitting and construct the psychological barycentric line (or most satisfactory local solution line) $L(b(\{f^h(z)\}))$; and then calculate the global satisfaction degree according to (8), and sort them in descending order to obtain the most satisfactory solution.

$$sat(f, Z) = \left(\frac{\sum\limits_{i=1}^{m} f_{ij}}{\sum\limits_{j=1}^{m} F_j}\right)^{\left(\frac{\sum\limits_{i=1}^{m} f_j}{3(\sum\limits_{j=1}^{m} f_{ij})}\right)} * \exp\left(-\frac{\sum\limits_{j=1}^{m} w_j |f_j - b(f^h(z_j)|}{\sum\limits_{j=1}^{m} w_j \delta_j}\right) \tag{8}$$

Where, $sat(f, Z)$ is the satisfaction of evaluated object f from evaluator Z, whose value is expected to be between 0 and 1. f_j is the value of each indicator. $|f_j - b(f^h(z_j)|$ is to measure the difference between each attribute value and the corresponding barycentric value. w_j and δ_j are used as the factor which can be adjusted to make the satisfaction comparable value in the case where the original results are not desirable. $\sum\limits_{j=1}^{m} F_j$ is the sum of F_j with each indicator value full score. $\sum\limits_{ij=1}^{m} f_{ij}$ is the sum of the values of all the indicators F_{ij} of F_i.

4   Simulation Experiment

To verify the effectiveness of the improved method, we chose the grades of nine courses from 2008 students in the final exam in a high school as the experimental data, nine courses being taken as nine indicators including Chinese, mathematics, English, physics, chemistry, politics, history, geography and biology. The sample data is shown in Table 1.

First of all, we use the attribute coordinate comprehensive evaluation method to respectively construct the psychological barycentric curves of several courses without applying the principal component analysis. And then we improve the method in the

Table 1. Sample data of nine courses

Item	Chinese	Math	English	Physics	Chemistry	Politics	History	Geography	Biology
1	91	68	82	27	55	72	78	71.5	75.5
2	91	77	50	53	65	47	75	70.5	80.5
3	91	88	15	78	48	65	63	66.5	79

way that the principal component analysis is used to simplify the indicators, further the attribute coordinate comprehensive evaluation method is applied to construct the psychological barycentric lines of the new indicators.

We also compare the global satisfaction degrees between two students before and after the improved method is applied.

4.1 Attribute Coordinate Comprehensive Evaluation Without Using Principal Component Analysis

Respectively, we choose the total score of 1000, 701 and 620 as the three evaluation planes, and select some samples for the experts to rate. The last column (Rating) of Tables 2 and 3 are respectively the rating data for total score 701 and 620.

Table 2. The samples and ratings for total score 701

Item	Chinese	Math	English	Physics	Chemistry	Politics	History	Geography	Biology	Rating
1004	97	96	71	75	68	55	78.5	71.5	90.5	8
1005	89	86	66	80	81	69	74.5	69	87	9
1476	94	102	48	74	76	60	77.5	82.5	87	9
398	69	84	80	82	74	70	75	81	85	6

Table 3. The samples and ratings for total score 620

Item	Chinese	Math	English	Physics	Chemistry	Politics	History	Geography	Biology	Rating
1345	85	95	61.5	71	54	47	68	69.5	70	7
1	91	68	82	27	55	72	78	71.5	75.5	9
947	95	90	72.5	79	35	55	65.5	61	67	10
1023	86	80	72.5	52	63	51	76	59	80.5	9
1548	85	89	67.5	69	41	41	71.5	79	77	8

According to (7), the barycentric coordinates of total score 701 and 620 with (Chinese, math, geography) are respectively (88.65625, 92.625, 76.4375) and (88.79069767, 83.93023256, 67.51162791).

Next, according to the interpolation theorem, we calculate the barycentric curves of Chinese, mathematics, geography (respectively shown in Figs. 1, 2, 3). It can be seen that the barycenter curve of Chinese is very unreasonable, as the curve should be monotonically increasing, while in this curve, the curve for total score of 650 is even lower than that of the total score of 600. From Figs. 2 and 3, we can see that barycentric curves of mathematics and geography are almost the same, which is not obvious to see the expert put more weight on arts or science.

The most likely reason for the result is that so many indicators make it difficult for experts to accurately distinguish good samples from bad samples among nine indicators, which could result in arbitrary ratings.
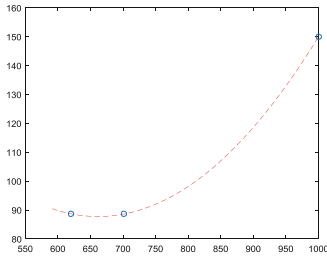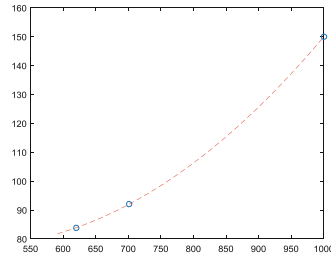

Fig. 1. The barycenter curve of Chinese
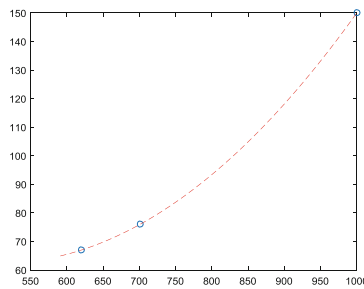

Fig. 2. The barycenter curve of Math

Fig. 3. The barycenter curve of Geography

4.2 Attribute Coordinate Comprehensive Evaluation with Principal Component Analysis

We apply the improved algorithm, first carrying out principal component analysis to reduce the quantity of indicators.

(1) **Calculate the covariance matrix S (correlation coefficient matrix) between indicators.**

$$
\begin{pmatrix}
 & x1 & x2 & x3 & x4 & x5 & x6 & x7 & x8 & x9 \\
x1 & 0.1725 & 0.3276 & -0.2918 & 0.5982 & 0.0915 & -0.3520 & -0.3481 & 0.2933 & -0.2836 \\
x2 & 0.5151 & -0.3978 & 0.5263 & 0.4750 & 0.0890 & 0.2451 & -0.0012 & -0.0697 & -0.0318 \\
x3 & 0.2543 & 0.7953 & 0.4751 & -0.1837 & -0.0851 & 0.1512 & 0.0638 & 0.0963 & 0.0034 \\
x4 & 0.4274 & -0.1859 & 0.0203 & -0.2104 & -0.6264 & -0.5471 & 0.0743 & 0.1566 & 0.1282 \\
x5 & 0.3474 & -0.0728 & -0.0053 & -0.3652 & 0.7509 & -0.3622 & 0.1624 & 0.1384 & -0.0019 \\
x6 & 0.1596 & 0.1882 & -0.2401 & 0.2653 & 0.0964 & -0.0190 & 0.0561 & -0.2374 & 0.8614 \\
x7 & 0.2310 & 0.1357 & -0.3541 & 0.1628 & -0.0799 & 0.0730 & 0.7306 & -0.3300 & -0.3492 \\
x8 & 0.3152 & -0.0823 & -0.4071 & -0.1534 & -0.0618 & 0.5785 & -0.0036 & 0.5999 & 0.0786 \\
x9 & 0.3983 & 0.0389 & -0.2510 & -0.2912 & -0.0266 & 0.1494 & -0.5532 & -0.5754 & -0.1788
\end{pmatrix}
$$

(2) **Calculate the eigenvalue vector of the correlation coefficient matrix**

(1.5315, 0.2945, 0.2291, 0.1658, 0.1331, 0.1170, 0.1006, 0.0881, 0.0778)

(3) **Calculate the principal component contribution rate vector λ and cumulative contribution rate G(M).**

The contribution rate vector λ = (55.9456, 10.7591, 8.3684, 6.0553, 4.8631, 4.2733, 3.6763, 3.2172, 2.8418)

The contribution rate of the first three principal components is G(M) = 75.0731%, although there will be some information loss, it is not so great to affect the overall situation.

According to the coefficient matrix S, the expressions of the first three principal components (f1, f2, f3) are respectively as follows.

f1 = 0.1725x1 + 0.5151x2 + 0.2543x3 + 0.4274x4 + 0.3474x5 + 0.1596x6 +
0.231x7 + 0.3152x8 + 0.3983x9
f2 = 0.3276x1 − 0.3978x2 + 0.7953x3 − 0.1859x4 − 0.0728x5 + 0.1882x6 +
0.1357x7 − 0.0823x8 + 0.0389x9
f3 = −0.2918x1 + 0.5263x2 + 0.4751x3 + 0.0203x4 − 0.0053x5 − 0.2401x6 −
0.3541x7 − 0.4071x8 − 0.2510x9

Respectively, x1, x2 … x9 represents Chinese, math…biological.

From the expression of the first principal component f1, it has the positive load on each variable, indicating that the first principal component represents the comprehensive components.

From the expression of the second principal component f2, the value of f2 decreases with the increase of x2(Math), x4(physics) and x5(chemistry), whereas increases with the increase of x3(English), x6(politics), x7(history) and x9(biology), which indicates f2 reflects a student's level of liberal arts.

From the expression of the third principal component f3, the value of f3 increases with the increase of x2(Math), x3(English) and x4(physics), whereas decreases with the increase of x1(Chinese), x6(politics), x7(history), x8(geography) and x9(biology), which indicates f3 reflects a student's level of science.

In this way we can simplify the nine indicators into three ones: f1, f2 and f3. Now we can calculate students' scores with the new indicator system. Table 4 is new sample data with the new indicator system.

Table 4. Sample data with the new indicators

Item	f1 (comprehensive)	f2 (liberal arts)	f3 (science)
1	184.3496	80.13484	55.48093
2	192.4282	40.70861	51.70827
3	192.2151	7.118305	43.40139

(4) **Attribute coordinate comprehensive evaluation**

   Respectively we provide three total score planes 460, 345 and 311 for the expert to rate. The scores of the last two total samples are shown in Tables 5 and 6 respectively. The expert's preference can be seen directly from the ratings (the last column). When the total score is higher, the expert pays more attention to the comprehensive level of students. When the total score is relatively lower, the expert values students' science scores more. This evaluation is easier than that without principal component analysis.

Table 5. The samples and ratings for total score around 345

Item	f1 (comprehension)	f2 (liberal arts)	f3 (science)	Total score	Ratings
263	238.4894	44.51444	62.95037	345.9542	7
999	228.5096	62.76003	54.17796	345.4476	6
1007	217.0385	52.21342	75.05868	344.3106	10
1066	245.756	37.56391	60.98198	344.3019	9

Table 6. The samples and ratings for total score around 311

Item	f1 (comprehension)	f2 (liberal arts)	f3 (science)	Total score	Ratings
1074	224.2676	29.82561	57.03609	311.1293	6
1699	166.3697	68.02551	76.22716	310.6224	10
798	205.9612	43.66174	60.83696	310.4599	8
735	199.5349	43.14419	67.52119	310.2003	9

   We can obtain the barycentric coordinates of 460, 345 and 311 respectively (268.2157, 92.1146, 100), (223.98, 51.46642, 6922313) and (200.2936, 44.12158, 66.05498). We draw the barycentric curves of indicator f1, f2 and f3 respectively (shown in Figs. 4, 5, 6). It can be seen that the three curves are all monotonically increasing, which are more reasonable than those drawn with the old model.
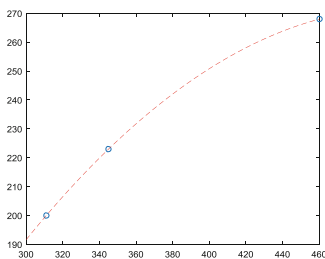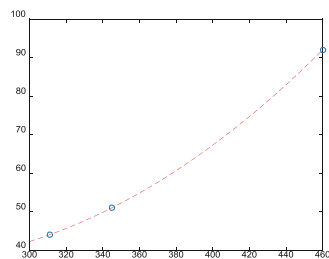

Fig. 4. The barycentric curve of indicator f1

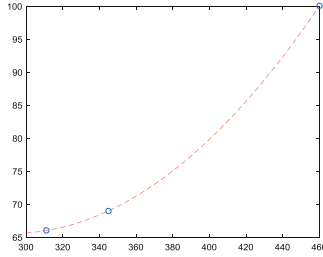**Fig. 5.** The barycentric curve of indicator f2

Fig. 6. The barycentric curve of indicator f3

(5) **the Comparison of Satisfaction Degree Before and After Improvement**

Finally, we examine the satisfaction degree obtained respectively using the two models. The followings are the scores of two students No. 466 and No. 196. They almost have the same total score, however, it is obvious that No. 196 is better at science than No. 466. So normally the satisfaction degree of No. 196 should be greater than that of No. 466 under the condition that the evaluator values the science scores more. However the result is opposite in the case of the unimproved method, which is unreasonable (shown in Table 7). Comparatively, the improved algorithm fixes the flaw and obtains the reasonable result, better reflecting the preference of the evaluator (shown in Table 8).

Table 7. The comparison of satisfaction degrees using the unimproved method

Item	Chinese	Math	English	Physics	Chemistry	Politics	History	Geography	Biology	Satisfaction
466	98	65	80	77	53	64	75	55	79.5	0.7896
196	80	72	31	81	71	62	82	80.5	86	0.7745

Table 8. The comparison of satisfaction degrees using the improved method

Item	f1 (comprehension)	f2 (liberal arts)	f3 (science)	Satisfaction
196	255.5199	93.5100	126.7342	0.5708
466	220.5927	118.4874	110.3656	0.5524

5   Conclusion

The improved method integrates principal component analysis into the original method to reduce the number of indicators so as to make the experts' rating process more simple and effective. The simulation examines the comparison of the results before and after using the principal component analysis and shows that the barycentric curves look more favorable, and the satisfaction degrees of the evaluated objects more accurately reflect the preferences and experiences of experts.

References

1. Ji, Z.: Research on personal information security evaluation based on analytic hierarchy process. Value Eng. **5**, 57–60 (2018)
2. Gong, X., Chang, X.: Comprehensive prejudgment of coal mine fire rescue path optimization based on fault tree-analytic hierarchy process. Mod. Electron. Tech. **8**, 151–154 (2018)
3. Pan, X., Liu, F.: Evaluation mode of risk investment based on principal component analysis. Sci. Technol. Prog. Policy **3**, 65–67 (2004)
4. Zhou, B., Wang, M.: A method of cloud manufacturing service QoS evaluation based on PCA. Manuf. Autom. **14**, 28–33 (2013)
5. Xu, X., Xu, G., Feng, J.: Study on updating algorithm of attribute coordinate evaluation model. In: Huang, D.-S., Hussain, A., Han, K., Gromiha, M.M. (eds.) ICIC 2017. LNCS (LNAI), vol. 10363, pp. 653–662. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63315-2_57
6. Xu, G., Xu, X.: Study on evaluation model of attribute barycentric coordinates. Int. J. Grid Distrib. Comput. **9**(9), 115–128 (2016)
7. Xu, X., Feng, J.: A quantification method of qualitative indices based on inverse conversion degree functions. In: Enterprise Systems Conference, pp. 261–264 (2014)
8. Xu, G., Min, S.: Research on multi-agent comprehensive evaluation model based on attribute coordinate. In: IEEE International Conference on Granular Computing (GrC), pp. 556–562 (2012)
9. Xu, X., Xu, G., Feng, J.: A kind of synthetic evaluation method based on the attribute computing network. In: IEEE International Conference on Granular Computing (GrC), pp. 644–647 (2009)
10. Xu, X., Xu, G.: Research on ranking model based on multi-user attribute comprehensive evaluation method. In: Applied Mechanics and Materials, pp. 644–650 (2014)
11. Xu, X., Xu, G.: A recommendation ranking model based on credit. In: IEEE International Conference on Granular Computing (GrC), pp. 569–572 (2012)
12. Xu, X., Feng, J.: Research and implementation of image encryption algorithm based on zigzag transformation and inner product polarization vector. In: IEEE International Conference on Granular Computing, vol. 95, no. 1, pp. 556–561 (2010)
13. Xu, G., Wang, L.: Evaluation of aberrant methylation gene forecasting tumor risk value in attribute theory. J. Basic Sci. Eng. **16**(2), 234–238 (2008)