Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 2, No. 1, 2024

Design and Optimisation of Research Performance Allocation Based on Principal Component Analysis

Shengzhen Ding[1,*], Xuanyu Chen[1], Weipeng Chen[2], Shaofeng Chen[2]

[1]School of mathematics and statistics, Zhaoqing University, 526061, Guangdong, China
[2]School of Machinery and Automobile Engineering, Zhaoqing University, 526061, Guangdong, China
[*] Corresponding author

Abstract: In the field of scientific research, scientific and reasonable scientific research performance allocation can mobilise the enthusiasm of researchers to the greatest extent, which is conducive to promoting the effective transformation and vigorous development of science and technology. In this paper, for the problem of scientific research performance allocation, the model of principal component analysis is established to be used for programme design, and the 2023 scientific research achievement incentives of 20 scientific research post employees are evaluated. A reasonable scheme is formulated to determine the priority of bonus distribution by the ranking of scientific research achievements and to distribute the bonus by using the law of two-eighths.

Keywords: Principal Component Analysis (PCA); Research Performance Allocation; Variance Contribution Ratio; Data Normalisation; Bonus Allocation Scheme.

1. Introduction

In the field of scientific research, a scientific and reasonable performance bonus allocation scheme is crucial for mobilising the enthusiasm of researchers, promoting the transformation of scientific and technological achievements as well as the development of the research field. Traditional performance allocation methods are often difficult to reflect the actual contributions of researchers fairly and accurately, therefore, this study proposes a performance allocation design and optimisation scheme based on principal component analysis (PCA) [1]. Principal component analysis is a statistical technique that extracts the most important information by reducing the dimensionality of the data, and this method is particularly suitable for dealing with complex datasets containing multiple variables. In this study, by building an evaluation model centred on PCA, we provide a comprehensive and objective assessment of researchers' performance [2]. The method not only reveals the major trends and patterns in the data, but also provides data support for the rational allocation of research performance. In addition, we have introduced the law of two-eight to optimise the allocation of bonuses, ensuring that outstanding researchers are given appropriate incentives, and that the entire allocation process is fair and transparent [3]. This data-driven performance allocation scheme can better stimulate the innovation potential and enthusiasm of researchers, thus promoting scientific research to a higher level.

2. Related Work

In research on research performance allocation, related work has focused on exploring how to effectively use various statistical and mathematical methods to evaluate and optimise the allocation of performance bonuses. In this regard, Principal Component Analysis (PCA) provides us with a powerful tool that captures the most important variables in the data through dimensionality reduction, thus helping us to understand more accurately the contribution and performance of researchers. In addition, the entropy weighting method is also used to evaluate the performance of researchers, which determines the weights by calculating the information entropy of each indicator, further enhancing the objectivity and fairness of performance assessment [4].

In conducting performance allocation, this study not only refers to previous work, such as using PCA and entropy weight method to quantify the contribution of researchers, but also innovatively combines the law of two-eight to design the bonus allocation scheme. This combination takes into account the non-linear characteristics of performance and attempts to motivate the overall motivation and efficiency of the research team by allocating most of the bonus to the best performing researchers.

Through these methods, we are able to establish a more fine-grained and dynamic performance evaluation system, which not only helps research managers to allocate resources more rationally, but also stimulates the innovation potential and research enthusiasm of researchers [5]. In addition, the application of these methods also promotes the theoretical development in the field of performance management and provides new perspectives and tools for future research performance management. Based on the synthesis of these related works, the study in this paper not only proposes new solutions to the problem of research performance allocation, but also makes useful contributions to both the practice and theory of research performance management.

3. Principal Component Analysis (PCA)

Principal component analysis aims to explain as much information as possible about the original data in the least number of dimensions by linearly combining multiple indicators with certain correlations. The aim is to achieve a dimensionality reduction of the data such that the reduced dimensionality variables are linearly independent of each other and the new variables are linear combinations of the original variables [6]. As the number of principal components

increases, the proportion of the total variance explained by each principal component decreases, and thus the ability to synthesise the original information diminishes.

Principal component analysis helps us to better understand the structure and patterns of data by extracting key features from the data, while simplifying the data set and retaining as much information as possible. By interpreting and applying principal component analysis appropriately, we can deal with complex datasets more efficiently and discover underlying patterns and associations, providing strong support for further data analysis and decision-making [7]. This question belongs to the evaluation category, which involves more relevant elemental indicators, and principal component analysis is used to ensure objectivity.

In this paper, we use python code to calculate the variance contribution rate of each component in the dataset, through the contribution rate of principal components to the interpretation of the variable (which can be understood as exactly how many principal components are needed to express the variable as 100%). In general, the higher the variance explained ratio, the more important the principal component is, and the higher the weight share should be. Based on the standardised dataset the covariance matrix R is calculated.

$$R = (r_j)_{n\times n} = \begin{bmatrix} r_{11} & r_{12} & \cdots r_{1n} \\ r_{21} & r_{22} & \cdots r_{2n} \\ \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & r_{nn} \end{bmatrix} \quad (1)$$

Calculate the eigenvalues of matrix R into $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ and the corresponding eigenvectors $u_1, u_2 \ldots u_n$,where $u_j = (u_{1j}, u_{2j}, \cdots, u_{nj}), u_{nj}$ denotes the nth component of the jth eigenvector; the eigenvectors are composed of n new indicator variables.

$$\begin{aligned} y_1 &= u_{11}\bar{x}_1 + u_{21}\bar{x}_2 + \ldots + u_{n1}\bar{x}_n \\ y_2 &= u_{12}\bar{x}_1 + u_{22}\bar{x}_2 + \ldots + u_{n2}\bar{x}_n \\ y_n &= u_{1w}\bar{x}_1 + u_{2n}\bar{x}_x + \ldots + u_{nn}\bar{x}_n \end{aligned} \quad (2)$$

Where, y_1 is the 1st principal component, y_2 is the 2nd principal component,…, y_n is the nth principal component. Calculate the contribution of each principal component $b\ j(j = 1,2.,n)$ and the cumulative contribution of $y_1, y_2, \cdots, y_n (p \leqslant n)$.

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{n} \lambda_k} (j = 1,2, \ldots, n) \quad (3)$$

$$\alpha_p = \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{n} \lambda_k} (j = 1,2, \ldots, n) \quad (4)$$

After standardising the data, we standardise the data for principal component analysis (PCA) to determine the contribution of each indicator to the principal component through the proportion of variance of each component, and the variance interpretation table in Table 1 below shows that the cumulative contribution rate of the first eight principal components reaches more than 90%, so the eight principal components are selected as the core indicators of the next data processing.

Table 1. Explanation of variance

Component	Eigenroot	Variance explained (%)	Cumulative variance explained (%)
1	4.822	30.138	30.138
2	2.485	15.534	45.673
3	2.053	12.829	58.501
4	1.504	9.401	67.902
5	1.222	7.64	75.542
6	1.088	6.799	82.342
7	0.825	5.156	87.497
8	0.681	4.253	91.751
9	0.511	3.193	94.944
10	0.318	1.988	96.932
11	0.224	1.403	98.335
12	0.091	0.572	98.906
13	0.087	0.541	99.447
14	0.058	0.361	99.808
15	0.021	0.131	99.939
16	0.01	0.061	100

In this paper, after obtaining the cumulative contribution ratio of variance, we use python code to score the dataset after Z-score normalisation, and sort the results according to the score and save them to the evaluation results table.

For the allocation scheme of bonus distribution, first, we determined the percentage of fixed bonus addition. It was decided that 20% of the performance bonus pool would be used for the fixed bonus add-on, so this would be 20% of $500,000, i.e., $100,000. The remaining 80% of the performance bonus pool is 80% of $500,000, or $400,000 [8]. The next step is to distribute the $400,000 in accordance with the rule of two-eight. Using the Rule of Twenty-Eight, we divide this amount into an 80% main portion and a 20% tail portion. The main portion is 80% of the $400,000, or $320,000, and the tail portion is 80% of the $400,000, or $320,000. The tail portion is 20% of $400,000, or $80,000. Now that we have determined the amounts of each component, we have a fixed bonus of $100,000, a main component of $320,000, and a tail component of $80,000.

The next step is to distribute the bonuses according to the performance and ranking of the specific employees in accordance with the previously set distribution ratio, which is based on the rule of two-eight to distribute the bonuses of the main part and the tail part [9]. In addition, to ensure that the allocation process is fair, transparent and matches the performance of the employees to incentivise their positive

work performance.

When designing a performance bonus allocation programme, there are a number of aspects that need to be considered:

Fairness: This scheme ensures fairness as it divides the bonus into a fixed component and a performance-based component. The fixed portion is allocated independently of individual performance, while the performance-based portion is allocated based on the employee's performance, which helps to ensure that the bonus is distributed fairly and gives higher rewards to employees who perform well [10].

Motivational: By using the rule of two-eight and dividing the bonus into a main part and a tail part, employees are incentivised to strive for higher performance. The main portion takes up the lion's share of the bonus pool, which means there is greater bonus potential for top performers, thus incentivising them to improve their performance.

Transparency: The allocation principles of this programme are clear and easy to understand, and the percentage of the bonus can be determined based on the ranking and performance of the employees. This increases the transparency of the distribution process, reduces employee concerns about unfair bonus distribution, and increases employee trust and satisfaction.

Implementability: The implementation of this programme is relatively simple as it adopts clear allocation principles and can be adjusted according to the actual situation. At the same time, the fixed bonus plus component can be used as a stable incentive mechanism, which helps to improve the stability and loyalty of employees [11].

Efficiency: This scheme can effectively utilise the resources of the performance bonus pool while ensuring a reasonable distribution of the bonus. By dividing the bonuses into different parts and distributing them according to the performance of the employees, the efficiency of the bonuses can be maximised and the long-term development of the company can be promoted.

Therefore, this allocation scheme is better in terms of fairness, incentives, transparency, enforceability and efficiency, and can be used as an effective performance bonus allocation scheme.

4. Entropy Weighting (Physics)

Model indicators based on entropy weight method, entropy value method according to the definition of information entropy, for a certain indicator, you can use the entropy value to judge the degree of dispersion of a certain indicator, the smaller the value of its information entropy, the greater the degree of dispersion of the indicator, the greater the impact of the indicator on the comprehensive evaluation (i.e., weight), if the value of a certain indicator are all equal, then the indicator does not play a role in the comprehensive evaluation.

The steps are:

The factors are normalised according to the number of each option:

For positive indicators:

$$\chi'_{ij} = \frac{\chi_{ij} - \min(\chi_{1j}, \chi_{nj}, \ldots, \chi_{nj})}{max(\chi_{1j}, \chi_{nj}, \ldots, \chi_{nj}) - min(\chi_{1j}, \chi_{nj}, \ldots, \chi_{nj})} \quad (5)$$

For negative indicators:

$$\chi'_{ij} = \frac{\chi_{ij} - \min(\chi_{1j}, \chi_{nj}, \ldots, \chi_{nj}) - \chi_{ij}}{max(\chi_{1j}, \chi_{nj}, \ldots, \chi_{nj}) - min(\chi_{1j}, \chi_{nj}, \ldots, \chi_{nj})} \quad (6)$$

Calculate the entropy of the Jth indicator:

$$e_j = -k \sum_{i=1}^{n} P_{ij} In(p_{ij}), j = 1, \cdots, m \quad (7)$$

where $k = 1/In(n) > 0$, satisfying $e_j \geq 0$; the information entropy redundancy (discrepancy) is calculated:

$$d_j = 1 - e_j, j = 1, \cdots, m \quad (8)$$

Calculate the weights of the indicators:

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j}, j = 1, \cdots, m \quad (9)$$

Composite scores were calculated for each sample:

$$S_i = \sum_{j=1}^{m} W_j X_{ij}, j = 1, \cdots, n \quad (10)$$

Where X_{ij} is the normalised data. According to the score of each influencing factor, the importance ranking of all factors can be obtained.

In this paper, Min-Max normalisation is carried out for the data, and python code is used to score each indicator data using entropy weight method, and finally the score of each member is obtained. The specific score of each member is shown in Table 2 below:

Table 2. Member-specific scores

Member No.	Score	Member No.	Score	Member No.	Score	Member No.	Score
1	42.14327	6	41.35457	11	25.73754	16	4.09912
2	44.5463	7	24.9485	12	10.81592	17	100
3	20.24785	8	10.39539	13	53.81518	18	25.47523
4	4.32982	9	43.55599	14	37.39643	19	17.33309
5	48.49111	10	33.65891	15	20.87215	20	30.54664

Individual members can be ranked by their scores, and for the distribution plan for bonus allocation, first, we determine the percentage of the fixed bonus plus. It is decided that 20% of the performance bonus pool will be used for the fixed bonus increase, so this will be 20% of $500,000, i.e., $100,000. The remaining 80% of the performance bonus pool is 80% of $500,000, or $400,000. The next step is to distribute the $400,000 in accordance with the rule of two-eight. Using the Rule of Twenty-Eight, we divide this amount into an 80% main portion and a 20% tail portion.

The main portion is 80% of the $400,000, or $320,000, and the tail portion is 80% of the $400,000, or $320,000. The tail

is 20 per cent of $400,000, i.e., $80,000. The amounts for each component have now been determined to be $100,000 for the fixed bonus plus, $320,000 for the main component, and $80,000 for the tail component. The next step is to distribute the bonuses according to the previously set allocation ratios based on the performance and ranking of the specific employees, and to distribute the bonuses for the main and tail portions in accordance with the rule of two-eighths. In addition, to ensure that the allocation process is fair, transparent and matches the performance of the employees, in order to incentivise their positive performance.

5. Conclusion

In this study, we designed and optimised the allocation scheme of research performance through Principal Component Analysis (PCA), aiming to motivate researchers more fairly and effectively, and to promote the transformation of scientific and technological achievements. By applying PCA to the researchers' performance data in FY2023, we successfully identified the most critical variables and evaluated the researchers' performance comprehensively accordingly. The analysis results show that the selection of the first eight principal components as the core indicators of performance evaluation can explain more than 90% of the data variance, ensuring the accuracy and comprehensiveness of the evaluation.

In addition, in conjunction with the law of two-eight for bonus allocation, we allocate 80% of the performance bonus pool according to the size of contribution, which not only improves the incentive effect of bonus allocation, but also maintains the fairness and transparency of the allocation. Specifically, we divide the bonus into two parts: fixed bonus plus and performance bonus, where the fixed bonus plus accounts for 20% of the total bonus to stabilise the basic income of the researchers; while the performance bonus part is dynamically adjusted according to the specific contribution of the researchers, which better reflects the difference in individual performance.

It has been proved that this bonus allocation scheme based on data analysis can effectively match the relationship between researchers' performance and rewards, which not only enhances the motivation of researchers, but also promotes the overall optimisation of the research environment. Through this method, we not only solve the problem of subjectivity in traditional performance evaluation, but also improve the transparency of evaluation and allocation, providing a reliable and feasible performance management tool for research institutions. Therefore, the research results in this paper not only have important practical significance for research management, but also provide valuable theoretical and methodological references for future research on research performance management.

6. Discussion

In this paper, we explore the effectiveness of using Principal Component Analysis (PCA) to design and optimise a research performance allocation scheme. By analysing data on the results of 20 researchers for the year 2023, we have successfully developed a performance allocation model that can maximise the motivation of researchers. Principal component analysis, as a powerful statistical tool, enables us to interpret a large amount of raw data information through fewer dimensions, so that we can accurately evaluate the contributions of researchers and allocate bonuses accordingly.

We first normalised the raw data, then calculated the covariance matrix and extracted the key eigenvalues and eigenvectors. These eigenvectors formed new indicator variables that helped us better understand the main trends in the dataset. The results of the analysis showed that the cumulative contribution of the first eight principal components exceeded 90%, which proved that the main components we selected were sufficiently representative of the vast majority of information in the overall data, thus providing a reliable basis for performance evaluation.

In addition, the law of two-eight was adopted for the distribution of performance bonuses. We allocated the majority of the performance bonus pool to the top-performing researchers, which not only fairly rewarded the top performers, but also motivated others to look up to them and improve the overall work efficiency and the quality of research results. Specifically, 20% of the bonus is used for a fixed bonus plus to ensure basic financial security for each researcher, while the remaining 80% is dynamically distributed according to the two-eight principle, which gives more prominence to the incentives for high-performance talents.

In summary, the design of research performance allocation based on PCA not only improves the scientificity and transparency of the allocation scheme, but also effectively promotes the motivation and innovation of researchers through a reasonable incentive mechanism. The application of this method may provide lessons and references for other scientific research institutions or organisations when dealing with similar issues, and help to promote the innovation and development of performance management practices within the scientific research field.

References

[1] Hou, G., Wang, J., & Fan, Y. (2024). Multistep short-term wind power forecasting model based on secondary decomposition, the kernel principal component analysis, an enhanced arithmetic optimization algorithm, and error correction. Energy, 286, 129640.

[2] Witanowski, Ł., Ziółkowski, P., Klonowicz, P., & Lampart, P. (2023). A hybrid approach to optimization of radial inflow turbine with principal component analysis. Energy, 272, 127064.

[3] Islam, S., Amin, S. H., & Wardley, L. J. (2024). A supplier selection & order allocation planning framework by integrating deep learning, principal component analysis, and optimization techniques. Expert Systems with Applications, 235, 121121.

[4] Ren, Q., Zhang, H., Zhang, D., & Zhao, X. (2023). Lithology identification using principal component analysis and particle swarm optimization fuzzy decision tree. Journal of Petroleum Science and Engineering, 220, 111233.

[5] Yesilkaya, B., Sayilgan, E., Yuce, Y. K., Perc, M., & Isler, Y. (2023). Principal component analysis and manifold learning techniques for the design of brain-computer interfaces based on steady-state visually evoked potentials. Journal of Computational Science, 68, 102000.

[6] Mezenner, N., Dechemi, N., Bermad, A., & Benkaci, T. (2024). Optimized reservoir operation using genetic algorithm and simulated inflows to reservoir based principal components analysis: case of Cheffia reservoir—Algeria. Modeling Earth Systems and Environment, 10(1), 383-391.

[7] Durkaya Kurtcan, B., & Ozcan, T. (2023). Predicting customer churn using grey wolf optimization-based support vector

machine with principal component analysis. Journal of Forecasting, 42(6), 1329-1340.

[8] Vinutha, M. R., Chandrika, J., Krishnan, B., & Kokatnoor, S. A. (2023). EPCA—enhanced principal component analysis for medical data dimensionality reduction. SN Computer Science, 4(3), 243.

[9] Zhong, C., Wu, P., Zhang, Q., & Ma, Z. (2023). Online prediction of network-level public transport demand based on principle component analysis. Communications in Transportation Research, 3, 100093.

[10] Cai, W., Deng, R., Gao, C., Wang, Y., Ning, W., Shu, B., & Chen, Z. (2023). Evaluation techniques for shale oil lithology and mineral composition based on principal component analysis optimized clustering algorithm. Processes, 11(3), 958.

[11] Zhang, X., Dai, X., Zhang, X., & Jin, G. (2023). Joint principal component analysis and total variation for infrared and visible image fusion. Infrared Physics & Technology, 128, 104523.