The Application of Data Mining inPayroll Distribution

Zhong Zheng[1]*, Lingli Huang[2], Tianlong Wang[3], Gang Wang[4], Wenjie Liu[5]

[1]College of Electrical Engineering & New Energy, China Three Gorges University, Yichang, 443002, China. [2]College of Science, China Three Gorges University, Yichang, 443002, China. [3]College of Civil Engineering & Architecture, China Three Gorges University, Yichang, 443002, China. [4]College of Mechanical & Power Engineering, China Three Gorges University, Yichang, 443002, China. [5]College of Foreign languages, China Three Gorges University, Yichang, 443002, China.

ABSTRACT

This paper studies the application of data mining in total wage distribution. The wage distribution model based on entropy method and analytic hierarchy process is established. Taking a state-owned enterprise as an example, the data was preprocessed with the linear equation fitting method. Entropy method was used to determine the weight of the influencing factors of wage distribution, and the first 8 factors were selected as the main influencing factors. The analytic hierarchy process (AHP) was used to calculate the weight of contract worker's salary and contract employees' salary as 0.342 and 0.658, respectively. On this basis, the total wage distribution. Compare and analyze the established distribution plan with the original distribution plan, and put forward improvement Suggestions to the original distribution plan: should increase the proportion of contract labor.

Keywords: The unitary linear equation fitting; Entropy method; AHP

1. Introduction

Total wage distribution is a management element closely related to human resource strategy. The total wage distribution mechanism of an enterprise is crucial to the development of the enterprise. It has an impact on employees' work enthusiasm, discipline compliance and order compliance, etc. Besides, the rational distribution of total wage helps the enterprise to achieve strategic goals, improve business performance, improve market competitiveness and strengthen corporate culture. How to establish a set of scientific, reasonable total wage distribution plan is a new and important task for state-owned enterprises. Based on an example of a state-owned enterprise, this paper establishes a salary distribution model based on entropy value method and analytic hierarchy process, and provides a reference for the improvement of salary distribution.

At present, the head office of a state-owned enterprise plans to distribute the total salary to the branches of 26 provinces and cities, that is, the state-owned enterprise plans to distribute a total of more than 36 billion yuan of wages to the branches of 26 provinces and cities in 2018, so it needs to ensure the reasonable distribution of wages to the branches of each province and city as far as possible. In order to allocate the total salary scientifically in the branches of 26 provinces and cities, promote the operation and development of the enterprises, improve the management level and enhance the competitiveness, the state-owned enterprise head office should take into account the regional differences, income, cost scale and income of each province and city.

In general, the head office are formulated at the beginning of this year the total wages of all branches scheme, at the end of this year's various branches based on the actual operation for fine-tuning, formulate the distribution plan for the next year, for example, in early 2018, according to the actual operations of the various provinces and cities in 2017 branch level allocation scheme in 2018, and implementation; At the end of 2018, according to the actual operation situation of the branches in each province and province in 2018, it will determine whether the total wage distribution formulated at the beginning of 2018 is reasonable, so as to fine-tune the distribution plan formulated in 2018 and formulate and implement the distribution plan in 2019 accordingly. For the branches in each province, there are contract workers and contract employees' workers, and the wage distribution of the two types of workers is also the result of comprehensive consideration.

Now we know the operation situation of each provincial and municipal company in 2018, and the total wage distribution plan formulated and implemented at the beginning of 2018.

2. Model assumptions

Hypothesis 1: There is little difference between the data collected in this paper and the actual value;

Hypothesis 2: The long-term operation of each branch is stable;

Hypothesis 3: The factors affecting the wage distribution discussed in this paper are all the influencing factors, which are not affected by other factors.

Hypothesis 4: The comprehensive development of each branch is stable, and emergencies are not discussed.

3. Problem analysis

In this paper, based on the actual operation situation of branches in various provinces and cities in 2018, the wage distribution model based on entropy value method and analytic hierarchy process is established, so as to realize the fine-tuning of wages and make the distribution of wages more reasonable. The idea diagram to solve the problem is as follows. First of all, the data should be preprocessed. Since there are too many data and too much data, the data should be cleaned in combination with real life.
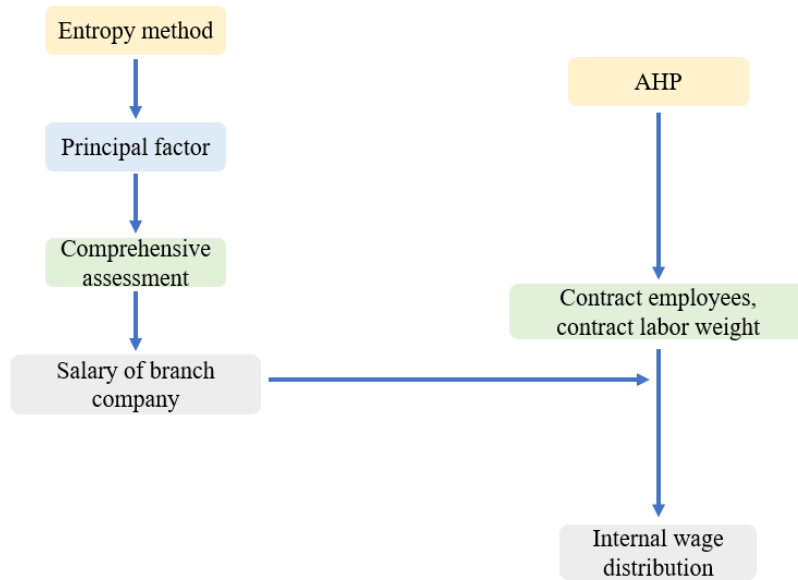
Fig.1: Idea diagram

Secondly, the entropy method is used to make a comprehensive ranking of all factors, find the main influencing factors, and combine the main influencing factors to make a comprehensive evaluation of each provincial and municipal companies, to determine the proportion of wage distribution; Then, for the wage distribution of different types of workers in the branch, the weights of each type of workers are determined by combining the analytic hierarchy process, and then the distribution is realized. Finally, the results of the new distribution plan are compared with the distribution plan formulated at the beginning of 2018 to achieve the fine-tuning of wages.

4. Data analysis and processing

Unitary linear regression fitting: the value of an economic index is often influenced by many factors. If only one of them is the main factor and plays a decisive role, then unitary linear regression fitting can be used for predictive analysis and relationship description.

4.1 Outlier processing

Based on the analysis of the data studied in this paper, it is found that the total business cost of Sichuan and Chongqing branches in each province in 2018 was recorded to be zero, which is obviously an outlier according to the actual life experience. There may be a certain relationship between total business cost and total business income. Therefore, through the data of branches in other provinces and cities and the data of previous years, MATLAB software can be used to fit and verify the relationship. The specific results are as follows:
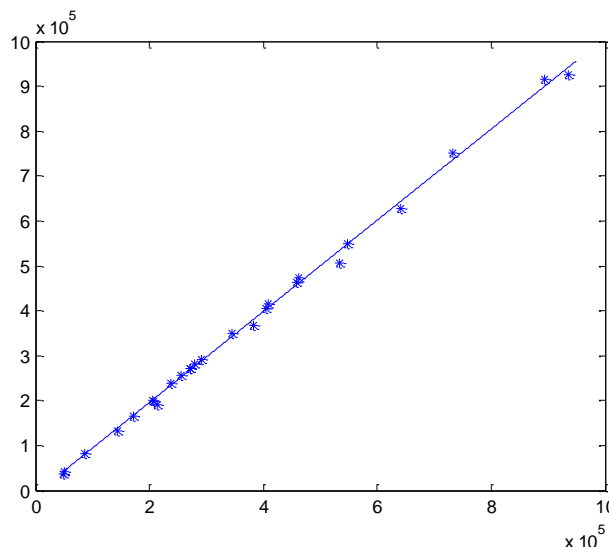

Fig.2: Total business cost and total business revenue fit

As can be seen from the figure, there is a high degree of fitting between the total revenue of business and the total cost of business, so the fitting curve can be obtained through fitting, and then the predicted value of the total cost of business can be obtained by bringing the corresponding total revenue of business into the fitting curve. The predicted value can be used as a reference for analysis.

4.2 Vacancy value processing

Through the analysis of the data studied in this paper, it is found that the regional population of Gansu is the missing value, which is unfavorable to the subsequent analysis in this paper. In this paper, combined with the actual situation, the regional population can be calculated by the urbanization rate and the number of urban residents, and the urbanization rate has a certain relationship with the income of urban residents, per capita disposable income of urban residents and GDP. Therefore, the population of the region can be calculated through the prediction of urbanization rate. Specific results are as follows:
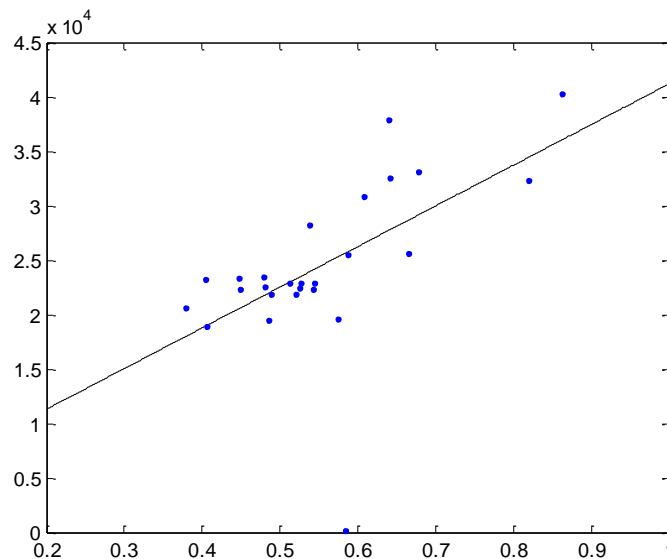

Fig.3: fitting results of urbanization rate and per capita disposable income of urban residents

The figure was obtained through clutch analysis of data by MATLAB. It is found that the fit degree is higher with the per capita disposable income of urban residents. According to the calculation, the corresponding missing value is obtained.

5. Model building

5.1 Theoretical basis

a. Entropy method: entropy method can be used to calculate the influence degree of indicators on comprehensive evaluation, that is, the importance degree of indicators;

b. AHP: the analytic hierarchy process (AHP) refers to the decomposition of the elements related to the decision into the levels of goals, criteria, plans, etc., on which the appropriate decision plan is determined.

5.2 Modeling

5.2.1 Comprehensive evaluation model based on entropy method

(1) Establish the data matrix:

Since the entropy method USES the ratio of a certain index of each scheme to the sum of the same index value, there is no influence of dimension. If there is a negative number in the data, the data needs to be de-negated. In addition, in order to avoid the meaninglessness of logarithms in entropy calculation, data translation is required.

$$A = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{mm} \end{pmatrix}$$

Type, X_{ij} represents the elements of matrix A after nonnegative transformation and data translation;

X_j represents the original matrix element;

X_{max}、X_{min} represents the largest and smallest elements per row.

The processed matrix A can be obtained. All the elements in A are in the interval [1,2].

$$x'_{ij} = \frac{x_j - x_{min}}{x_{max} - x_{min}}$$

The smaller the better indicator:

$$x'_{ij} = \frac{x_{max} - x_j}{x_{max} - x_{min}}$$

Type, x'_{ij} represents the data after non-negative processing;

x_{max} represents the maximum value of item j;

$$p_{ij} = \frac{x_{ij}}{\sum\limits_{i=1}^{n} x_{ij}} (j = 1, 2, \cdots m)$$

Type, x_{ij} represents the value of item j of branch i;

$\sum\limits_{i=1}^{n} x_{ij}$ represents the sum of data in

(2) Nonnegative processing of data

In the process of non-negativization of the data listed for a factor of that type, the data obtained from the non-negativization process is subtracted from the minimum or maximum value of the data listed for that factor, and then divided by the range of the data listed for that factor.

For the bigger is better factor:

x_{min} represents the minimum value of item j.

(3) Calculate the proportion of the ith branch in the index of item j

column j of factors;

p_{ij} represents the proportion of branch i in the index under item j;

(4) Calculate the entropy of the jth factor

$$e_j = -k * \sum_{i=1}^{n} P_{ij} \ln(P_{ij})$$

Type, k>0; the constant k is related to the number of samples m; general order $k = 1/\ln m$, $0 \le e \le 1$。

e_j represents the entropy of the jth

factor, which is generally greater than or equal to zero;

$\ln$ is the natural logarithm;

(5) Calculate the difference coefficient of factor j

$$g_j = 1 - e_j$$

Type, g_j represents the difference coefficient of factor j;

The larger the value, the more important the indicator. For item j, the greater the difference of

index value X_{ij}, the greater the effect on scheme evaluation, and the smaller the entropy value.

(6) Weight

$$W_j = \frac{g_j}{\sum\limits_{j=1}^{m} g_j}, j = 1, 2 \cdots m$$

Type, $\sum\limits_{j=1}^{m} g_j$ represents the sum of the difference coefficients of all factors;

W_j is the weight of the jth factor.

(7) comprehensive assessment

$$S_i = \sum_{j=1}^{m} W_j \cdot P_{ij} (i = 1, 2, \cdots n)$$

Type, S_i represents the comprehensive score of the branch in the ith province;

(8) Each province branch company income salary solution

Branch salary = The total wages×Comprehensive scores of all branches

5.2.2 Weight determination model based on analytic hierarchy process

(1) Establish hierarchy

AHP is used to solve practical problems by constructing a hierarchical structure. After defining the goal to be solved, it analyzes the factors that will affect the decision of the goal, and sorts out the relationships among the factors to form a coherent and hierarchical relationship chain. The hierarchical structure consists of three basic levels: target level, criterion level and measure level.

(2) Construct the judgment matrix and assign the value

a. Construct judgment matrix: Each element with a downward membership is taken as the first element of the judgment matrix (in the upper left corner), and the elements belonging to it are arranged in the first row and first column after it;

b. Fill in the judgment matrix: Compare the elements in pairs, and use 1 to 9 to indicate the importance of the elements to each other. The definition of scale is shown in table 2:

Table 1: Scale definition

Scale	Definition and description of two element comparisons
1	Two of equal (or equal) importance
3	This element is slightly more important (or stronger) than the other element
5	One element is more important (or stronger) than the other
7	One element is obviously more important than the other
9	One element is absolutely more important than another
2,4,6,8	A compromise scale between the two criteria

c. Fill in the judgment matrix: Compare the elements in pairs, and use 1 to 9 to indicate the importance of the elements to each other. Let the judgment matrix after filling in be:

$$A = \left(a_{ij} \right)_{n \times n}$$

A has properties:

$$a_{ij} = 1 \qquad a_{ji} = \frac{1}{a_{ij}} \qquad a_{ij} > 0$$

(3) Hierarchical single sort and check

a. Hierarchical single sort: Weight vectors are calculated by mathematical methods, including arithmetic mean method, geometric mean method, eigenvector method and least square method, etc. According to relevant literature, it is known that the least square method has a large error in the ranking results, while the other three have a small difference. In this paper, the eigenvector method is adopted to calculate the weight vector, and the specific formula is as follows:

$$M_i = \prod\nolimits_{j=1}^{n} a_{ij}$$

Type, M_i is expressed as the product of the elements in row i;

a_{ij} represents the importance ratio of the ith element and the jth element;

$$D_i = \sqrt[n]{M_i}$$

Type, D_i is the n-th square root of the product of the elements in row i;

M_i represents the product of the elements in row i;

$$\vec{D}_i = \frac{W_i}{\sum\limits_{i=1}^{n} D_i} \qquad \lambda_i = \sum\limits_{j=1}^{n} a_{ij} \vec{D}_j$$

Type, $\vec{D}_i$ represents the eigenvector;

λ_i Is the ith eigenroot, is the sum of

the product of $\vec{D}_i$ and a_{ij}, is equal to the sum of the elements times the corresponding eigenvectors;

$$\lambda_{\max} = \sum\limits_{i=1}^{n} \frac{\lambda_1}{n \times \vec{D}_i}$$

Type, $\lambda_{\max}$ is the maximum characteristic root, and n is the order of the judgment matrix.

b. Hierarchical single sort consistency test: In the process of constructing the judgment matrix, some subjective factors may lead to logical loopholes in the evaluation indicators, especially when the number of evaluation indicators is large. Therefore, it is necessary to verify the

consistency of the calculated results and prove that the obtained judgment matrix is a consistency judgment matrix, which is logically feasible. If it fails to meet the standard, the calibration value in the judgment matrix should be adjusted and recalculated.

Calculate the consistency index:

$$C.I. = \frac{\lambda_{\max} - n}{n - 1}$$

The above formula is used to determine the mean random consistency index $C.I.$ The value

of R.I. is shown in Table 3:

Table 2: The value of R.I.

n	1	2	3	4	5	6	7	8	9
R.I.	0	0	0.52	0.89	1.12	1.24	1.36	1.41	1.46

Calculate the consistency ratio:

$$C.R. = \frac{C.I.}{R.I.}$$

If $C.R. < 0.1$, the judgment matrix conforms to logic through consistency test; If $C.R. > 0.1$ is not logically valid, it should be corrected appropriately.

(4) Hierarchical general sort and check

a. Total hierarchical sorting: According to

the order from top to bottom, the relative weight of each factor in the final objective decision is calculated layer by layer. Suppose that the weight of the elements in the second layer relative to the target layer is:

$$A^{(n-1)} = \left(a_1^{(n-1)}, a_2^{(n-1)}, \cdots, a_h^{(n-1)} \right)^T$$

The relative weight of n total of k elements in layer n to any element j $(1 \le j \le h)$ in the

previous layer (layer $n-1$) is:

$$R_j^{(n)} = \left(R_{1j}^{(n)}, R_{2j}^{(n)}, \cdots, R_{kj}^{(n)} \right)^T$$

The weight of the elements independent of j is zero.

$$R^{(n)} = \left(R_1^{(n)}, R_2^{(n)}, \cdots, R_k^{(n)} \right)^T$$

Represents the weight vector set of the elements of layer n to layer $n-1$, then the weight of the elements of layer n to the total target is:

$$A^{(n)} = \left(a_1^{(n)}, a_2^{(n)}, \cdots, a_h^{(n)}\right)^T = R^{(n)} A^{(n-1)}$$

b. Consistency test of total hierarchical sorting: The total sort also requires a consistency check to verify the logical consistency of the overall hierarchical structure. Calculate the consistency indicator C.I. :

$$\text{C.I.}^{(n)} = \left(\text{C.I.}_1^{(n-1)}, \text{C.I.}_2^{(n-1)}, \cdots, \text{C.I.}_m^{(n-1)}\right) A^{(n-1)}$$

Determine the mean random consistency index C.R.:

$$\text{C.R.}^{(n)} = \frac{\text{C.I.}^{(n)}}{\text{R.I.}^{(n)}}$$

When $\text{C.R.}^n < 0$, the matrix is considered to conform to the consistency test.

(5) Wage distribution in branches of provinces and cities

Contract employees salary= Total salary of branch company×Contract employees wage weighting

Contract wages = Total salary of branch company×Wage weighting for contract workers

6. Model solution and result analysis

For each province branch company total salary reasonable allocation plan specific thinking is mainly divided into two steps, the first step, to the total salary distribution, according to the actual operation of the company. The comprehensive evaluation results of the first model can be analyzed. The second step is the distribution of wages in the branch based on the weight obtained through the analytic hierarchy process.

6.1 Major influencing factors

The following table is the table of partial data after forward standardization:

Table 3: Partial data after forward standardization

	1	2	3	4	5	6
1	1.153	2.000	2.000	1.223	1.645	2.00
2	1.089	1.516	1.581	1.134	1.170	1.800
3	1.671	1.142	1.124	1.469	1.236	1.559
4	1.303	1.139	1.088	1.235	1.181	1.556
5	1.191	1.259	1.315	1.171	1.125	1.631
6	1.379	1.330	1.319	1.380	1.286	1.633
7	1.216	1.169	1.180	1.175	1.113	1.551
8	1.324	1.166	1.179	1.277	1.159	1.484

Note: vertical lines: 1-8 are the serial Numbers of the cities attached to the no.1 middle school branch. Horizontal lines: 1-6 are the geographical population, per capita income of urban and rural residents, per capita consumption expenditure of urban and rural residents, number of urban residents, per capita income of urban residents and per capita disposable income of urban residents.

The data presented in the table is standardized, the data format and units are unified, and the size is limited from zero to one, making the comparison and processing of data more convenient.

Take the top eight factors of weight size as the main factors, as shown in the following table:

Table 4: Main factors and their weights

Indicators	Total revenue	Total operating cost	Regional population	Income of urban residents
The weight	0.062563534	0.062559577	0.062554499	0.062551566
Indicators	Other average wages	Fixed assets for production	Average selling price	Cost ratio
The weight	0.062539023	0.062525568	0.062516318	0.062514691

Note: Other average wages refer to the average wages of employees in other state-owned enterprises

6.2 Comprehensive evaluation of each branch

According to the main influencing factors selected, a comprehensive evaluation was conducted for each branch, and the following table (part) was obtained according to the evaluation results:

Table 5: Comprehensive evaluation results

City	Guangdong	Beijing	Jiangsu	Zhejiang
Evaluation result	0.04782	0.04536	0.04518	0.04313

Table 6: Rank of original salary distribution and rank of current salary distribution

	1	2	…	25	26
The original number	Guangdong	Jiangsu	…	Qinghai	Ningxia
Expected rank	Guangdong	Beijing	…	Guizhou	Hainan

It is obvious that the original plan ranked as Guangdong, Jiangsu, Zhejiang, Shandong, Ningxia, Guizhou and the south China sea, while Guangdong, Beijing, Jiangsu and Zhejiang were predicted to rank... Hainan, Qinghai, Ningxia. Obviously, it is not the same. Therefore, the total salary distribution of provincial and municipal branches set at the beginning of 2018 is not reasonable.

Comparison of new and old schemes: Compare the salary ratio of the branches in each province of the new and old schemes with the salary amount, as shown in the figure below:
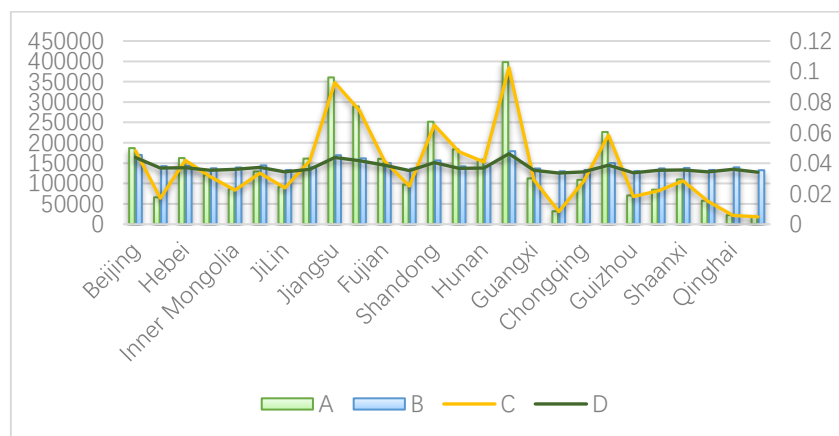

Fig.4: Comparison between the original allocation scheme and the new allocation scheme

In the figure, A, B, C and D respectively represent the proportion of the salary of each branch of the original plan, the proportion of the salary of each branch of the new plan, the salary of each branch of the original plan and the salary of each branch of the new plan. Through comparison, it can be found that in the original plan, the total salary of Jiangxi accounted for several times of that of Ningxia, Yunnan, Chongqing and other regions, and the wage distribution was extremely uneven and obviously unreasonable. Although there were differences in the new distribution plan, the overall difference was small, without too much fluctuation. Secondly, the higher total wage in the original plan is unreasonable in Guizhou and other places where transportation is inconvenient and the economic development level is low, while the higher total wage in the new plan is also more in line with the actual situation in Guangdong, Jiangsu and other areas with higher economic development level.

6.3 Wage distribution in branches of provinces and cities

Based on the practical problems studied in this paper, the analytic hierarchy process used in this paper is divided as follows:
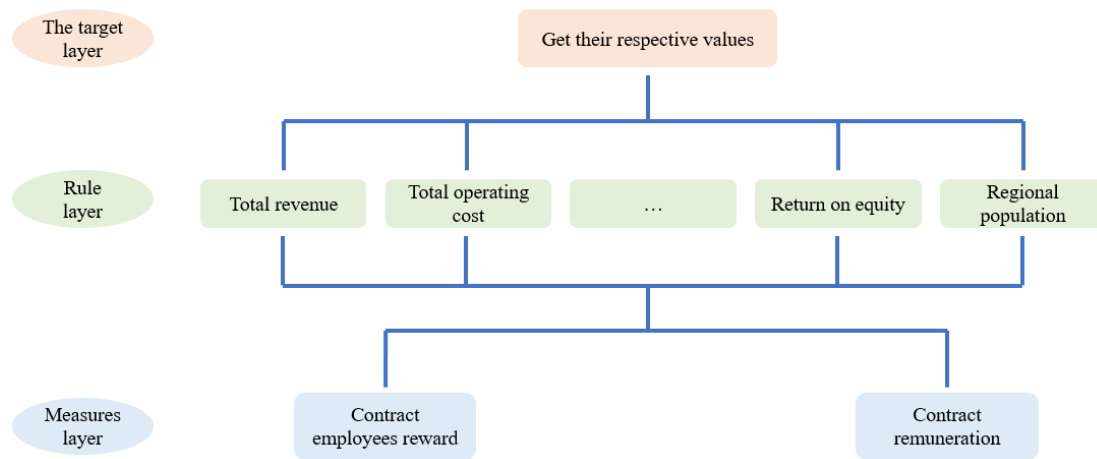

Fig.5: AHP hierarchy

The wages of contract workers are higher, but there is a limit to the number of contract workers in each branch; Contract employees, on the other hand, has lower salary and lower knowledge level, and more urban and rural residents in the city where the branch company is located. Therefore, the number of employees employed by the company can be increased. It is clear that when the company's operation is better and more money is made, the company's demand for contract employees is greater. To sum up, the factors considered in the distribution of labor remuneration of contract employees may be more related to the regional population, the income of urban and rural residents and so on. So it makes a better distinction between the two, and it makes a better contribution to the distributive weight.

(1) Judgment matrix and decision matrix
According to the research problem in this paper, the A matrix obtained is:

Table 7: A matrix

	A	B	C	D	E	F	G	H
A	1	2	2	5	3	3	4	4
B	1/2	1	2	4	2	3	3	4
C	1/2	1/2	1	3	2	3	3	2
D	1/5	1/4	1/3	1	2	2	1	3
E	1/3	1/2	1/2	1/2	1	1	2	3
F	1/3	1/3	1/3	1/2	1	1	2	3
G	1/4	1/3	1/3	1	1/2	1/2	1	1
H	1/4	1/4	1/2	1/3	1/3	1/3	1	1

Note: A~H are the main factors determined by the previous model

The horizontal and vertical axes in the table above are the order of the eight main factors obtained from the first question from high to low. It can be seen from the data in matrix A that the Numbers above the diagonal of the matrix are all greater than or equal to 1. The weights of the main factors obtained in the previous model are the results of the analysis.

Among them, the total revenue of business is larger than other factors in comparison value because of the largest weight, and so on. However, some judgments of importance do not

completely accord with the results of weights, because weights can only be used as one of the comparisons of importance, not all of them.

Therefore, also need to be flexible according to the actual situation.

$$\begin{bmatrix} 1 & 2 \\ 1/2 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 \\ 1/3 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 \\ 1/2 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 \\ 1/3 & 1 \end{bmatrix}$$

(a)   (b)   (c)   (d)

Fig.6: Partial decision level matrix

Note: The rows and columns of each matrix from left to right and from top to bottom are the labor remuneration of contract workers and contract employees.

(a) the criterion level factor considered is total business income, which may be more technical and more relevant to the contract workers. So contract workers are more important in this matrix.(b) the factor taken into account is the total cost of operations, which is more related to Labour and Labour, so Labour is more important in this matrix.(c) the factors taken into account are fixed assets for production, which are of greater importance because of their relative relation to the contract Labour. (d) the factor taken into account is the average selling price of commercial housing, which is more relevant to the contract workers, so the contract workers are more important. A similar matrix is established based on the understanding of the two kinds of rewards in the preparation of the model.

(2) Internal wage distribution results of branches in provinces and cities

The weights of contract employees and contract workers are 0.658 and 0.342, respectively. The total wage distribution and city ranking of each branch are as follows (part):

Table 8: Wage distribution within the branch

City	Contract employees	Contract worker	Aggregate amount
Ningxia	113856.0619	59268.5811	173124.6430
Qinghai	107985.2353	56212.4807	164197.7159
Gansu	107557.4464	55989.7921	163547.2385
Shaanxi	102687.8784	53454.9038	156142.7818
Yunnan	99465.7595	51777.6071	151243.3665
Guizhou	95359.3232	49639.9725	144999.2958
Sichuan	95341.4949	49630.6919	144972.1867
Chongqing	91790.5370	47782.2155	139572.7525

Note: Unit ten thousand yuan

In the two types of wage distribution within the branch, the labor remuneration of the urban branch listed in the table is greater than that of the contract workers.

The wages of contract workers are higher, but there is a limit to the number of contract workers in each branch; Contract employees, on the other hand, has lower salary and lower know-ledge level, and more urban and rural residents in the city where the branch company is located. Therefore, the number of employees employed by the company can be increased. It is clear that when the company's operation is better and more money is made, the company's demand for contract employees is greater.
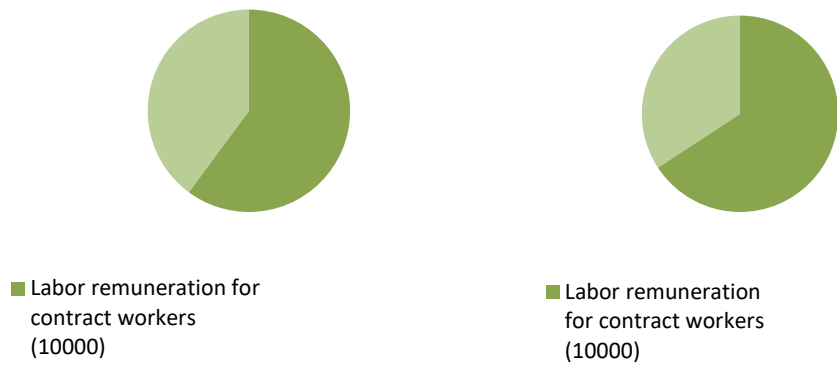
Fig.7: The ratio of the total salary of the original plan (left) and the new plan (right)

In the original plan and the new plan, the ratio of contract employees to contract remuneration is 0.67 and 0.51 respectively. (1) Root cause: the actual operating conditions of each branch changed in 2018 compared with the previous year; (2) Immediate cause: the weight factor considered the distribution of the two rewards has changed. Pie chart analysis can analyze the overall decrease in employment demand for the labor force by 2018. Therefore, overall should increase the proportion of contract workers.

7. Model evaluation

7.1 Advantages of the model

(1) There are many factors that need to be considered in the distribution of total wages, which is a comprehensive system. As one of the research methods of comprehensive indicators, entropy method makes the calculation results more objective and scientific because it eliminates the influence of human subjective factors. It can be used here to weight the influencing factors and obtain the comprehensive value；

(2) Analytic hierarchy process (AHP) is a method to transform multi-element decision problem into quantitative calculation problem. AHP is a flexible, systematic and systematic method, which can effectively solve the problem of multi-objective evaluation；

(3) In this paper, we preprocess the data and apply the method of unitary linear regression fitting to make the data more reasonable.

7.2 Model shortcomings

When judging the importance of each factor, although it is by ranking the importance of weights, the test matrix is by artificial estimation, which makes it subjective to some extent.

8. Improvement and promotion of the model
8.1 Model improvement

For a large number of influencing factors, we can also adopt the idea of dimensionality reduction, which can be realized by SPSS software.8.2 Generalization of the model

8.2 Generalization of the model

This paper studies the problem of data mining and wage distribution, which requires reasonable processing of data and comprehensive consideration of the influence of various factors. As one of the research methods of comprehensive index, the entropy method is objective and scientific, which is not only applicable to the problem of wage distribution, but also to the similar comprehensive evaluation and prediction of various kinds of water quality judgment.

9.Conclusion

This paper establishes a salary distribution model based on entropy method and analytic hierarchy process. Taking a state-owned enterprise as an example, eight factors, including Total revenue, Total operating cost and Regional population, were identified as the main factors affecting wage distribution. The weights of contract wages and contract employees wages are 0.342 and 0.658, respectively. The total salary of contract workers should be increased, and the distribution plan for 2018 has been improved according to the actual operation. For example, the increase of 343,977,525 yuan in Chongqing and the decrease of 73,653 yuan in Sichuan. The whole model can also be applied to practical problems such as task assignment and can provide reference for solving practical problems.

References

1. Lina Sun, Kai Wang. Comprehensive evaluation and spatial analysis of county economy based on entropy method[J]. Journal of Liaoning university of technology (social science edition), 2019,21(02):21-24.

2. Qi Yu. Overview of data cleaning techniques in data mining[J]. Henan science and technology, 2018(20):21-23.

3. Yu Lu. Based on analytic hierarchy process (AHP), the weight of driving condition is determined[J]. Practical automotive technology, 2019(02):114-117.

4. Zhigang Jian, Xu Jin. Research and realization of data preprocessing in data mining[J]. Research in computer applications,2004(07):117-118+157.

5. Wei He, Siyuan Yang. New research on wage reform [J]. China labor science,1990(10):17-20.