School Choices in Fund Distribution

Le Wang

North China Electric Power University, Baoding 071000, China

644591488@qq.com

Abstract: This article deals with the issues of school choices in fund distribution. First of all, through comparing two data sets of candidate schools and non-candidate schools, we employ different degree comparison, discrete data statistics and other methods to find out the indicators which has exerted an influence on this kind of classification. Otherwise, principal component analysis is utilized to set up a comprehensive assessment model, rank the candidate schools and distribute the funds. Final results show that totally 29 sets of indicators are important indicators of classification. And through ranking candidate schools according to these indicators, the investment schools and fund distribution chart is obtained.

Keywords: School choices, principal component analysis, indicators

1.Introduction

The Goodgrant Foundation, a charitable organization, wants to help improve educational performance of undergraduates in the United States. To do this, the foundation intends to donate a total of $100,000,000 to an appropriate group of schools per year, for five years. We are required to develop a model to determine an optimal investment strategy. Its goal is to have the highest likelihood of producing a strong positive effect on student performance. At the same time, maximizing the function of the investment also should be taken into account. All in all, we must do so in a more reliable and scientific way instead of judging subjectively.

2. Index Selection

The difference in important indicators is the symbol of differences between the candidate school and other schools, this is also the basis of filtering a candidate school. After that, we believe the date list of candidate schools is absolutely not for no reason at all, it must be some significant characteristic to distinguish them from all the school database. So we contrast the two Table and attempt to find the significant difference. To some degree, the difference of these characteristic reflect some criteria how we determine the list of Candidate

School. What's more, these factors will be our important indicators to evaluate candidate schools.

Three steps were divided to screened indicators and important indicators have been successfully located.

2.1 Screening One

First, by using the method of similarity analysis to screen important index, we make mean contrast of 75 groups of variables between 2937 groups of candidate school data and data from 4868 non-candidate schools so as to filter variable data with relatively smaller similarity, thinking these data almost make no effect on determining the list of candidate school. The following is the visualization of mean contrast:
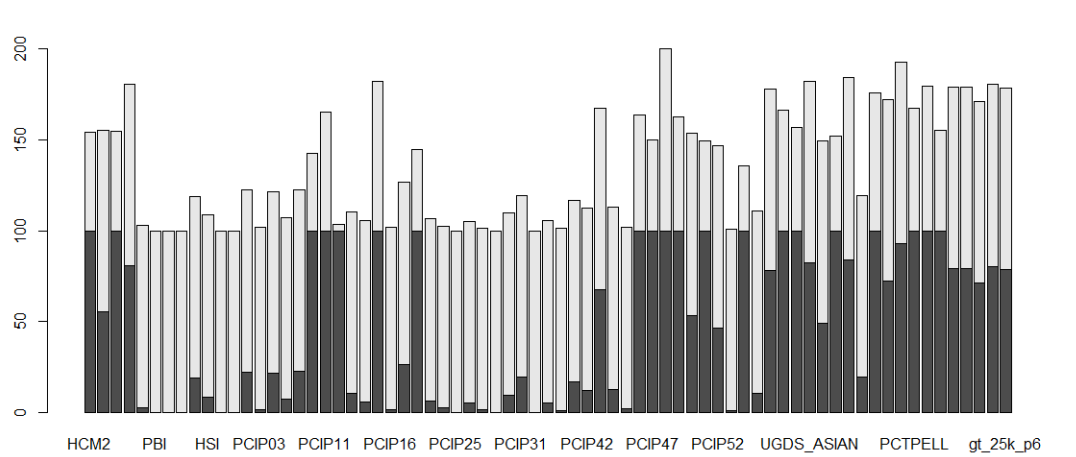

Fig 1. Visualization

To take the first step for screening, we establish a standard, introducing the conception of difference than, and the difference than is defined for each indicator as:

$$\Omega = \frac{|u - v|}{Max(u, v)}$$

Where u stands for the indicators of candidate schools

      v stands for the indicators of non-candidate schools

It reflects on similarity of two parameters, the smaller the greater the Ω is, the more similar they are. It is no doubt that the maximum of this parameter is 1 and the minimum of this parameter is 0.

For every index, calculate the Ω of candidate schools and non-candidate schools.

And evaluate it. The larger the Ω is, the lower the similarity of two groups of data index is, and the more discrete the distribution is. We make statistics for those indicators and establish a standard that for indicators whose difference than is higher than 50%, we think its influence no more than half and it's reasonable to give them up. After screening, the original 75 groups of indicators turn into 49 groups, which greatly reduces the number of indicators and is of great help to the results of evaluation. Model accuracy achieves further ascension now.

2.2 Screening Two

Though having deleted part of the indicators through the first step of screening, we still have 49 indicators left, which makes it still hard to evaluate schools. So other basis is needed to eliminate indicators of high similarity. This time, we introduce the

conception of DIFF to compare the differences of the contents of the two parameters, at the same time to eliminate the differences between different units. The greater the value of the Diff, the greater the degree of difference between two groups of data. Compare difference degree of candidate and the non-candidate schools' index under the same index, and give up indicators whose difference is not obvious.

To this end, we define a standard, that when the value of the diff is between average and 1, its difference degree is considered to be big, so the parameter can be preserved.

Through this step of filtering, we filter out nearly half of the indicators, the rest 29 set of indicators are in line with the normal data processing scale, and next we rank schools based on the 29 indexes.

Finally, it is found that the remaining indicators include not only the discrete index, but also the non-discrete index. We have a discussion on the two indicators.

```
> apply(data[,1:5],2,table)
  HBCU  PBI TRIBAL  HSI WOMENONLY
0 2841 2901   2902 2819      2893
1   95   35     34  117        43
> |
```

Fig 2. Discrete data statistics

For non-discrete indicators, the principal component analysis can be a good way to find the principal component and give the evaluation index, which is the necessary elements of the next school rankings.

They are:

Table 1. Important indicator

UGDS	PCIP09	PPTUG_EF	UGDS_ASIAN	cost	PCIP12	PCIP24
PCTFLOAN	UG25abv	UGDS_WHITE	UGDS_HISP	md_earn_wne_p10	gt_25k_p6	RPY_3YR_RT_SUPP
GRAD_DEBT_MDN_SUPP	GRAD_DEBT_MDN10YR_SUPP	UGDS_NRA	UGDS_UNKN	UGDS_2MOR		

3 Rank Schools

Make the main component analysis of the remaining 293719 group data, using Matlab software, we obtain the correlation coefficient matrix of the first 19 of the root and the contribution rate as shown in the table.

Table 2. Principal component analysis results

eigenvalues	contribution
22.4566	22.4566
7.6587	30.1153
6.4978	36.6130
6.1172	42.7302
5.4979	48.2281
5.2866	53.5147
5.2442	58.7589
5.2061	63.9650
4.9863	68.9513
4.4460	73.3973
4.3041	77.7014
3.9485	81.6499
3.2960	84.9459
3.1710	88.1168
2.8544	90.9713

It can be seen that the cumulative contribution rate of the former n characteristic root is more than 95%, and the principal component analysis is very good. The following is a comprehensive evaluation of the former n principal components.

The resulting *n* principal components are

The comprehensive evaluation model of principal components was constructed with the contribution rate of n principal components as the weight.

$$Z = 0.2246y_1 + 0.0766y_2 + 0.0650y_3 + 0.0612y_4 + 0.0550y_5$$
$$+0.529y_6 + 0.5241y_7 + 0.521y_8 + 0.0499y_9 + 0.0455y_{10}$$
$$+0.0430y_{11} + 0.0395y_{12} + 0.0330y_{13} + 0.0317y_{14} + 0.0285y_{15}$$

The *n* principal components into each year on the type, you can get the school ranking and the comprehensive evaluation results.

References

[1] Jeff Borland, Peter Dawkins, David Johnson and Ross Williams Returns to Investment in Higher Education [D] . Melbourne: The University of Melbourne

[2] Li xuan Xi, The Study of the Donation to American Higher Education[D]. Nanjing: Nanjing Normal University

[3] Zhe hui Zou, A Comparative Study in Higher Education foundation of Chi na and the United States[D]. Guangxi: Guangxi University

[4] Zhi hui Xu, The Method of Calculating the Markov State Transition Proba bility Matrix Research[D]. Heilongjiang: Northeast Agricultural University

[5] Xia Yang, Dongwei Wu, Application of R Language in large Data Processi ng[J]. Science and Technology Information, 2013,