Financial Aid Segmentation in Universities Based on CLS-RFM Model and Cluster Analysis

Zhongqiang Sun[1], Wenhao Ying[2(✉)], and Qing Xu[3]

[1] School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu, China
20205227120@stu.suda.edu.cn

[2] School of Computer Science and Engineering, Changshu Institute of Technology, Suzhou, Jiangsu, China
ywh@cslg.edu.cn

[3] College of Electrical and Automation Engineering, Changshu University of Technology, Suzhou, Jiangsu, China
xuqing@cslg.edu.cn

Abstract. At present, colleges and universities have established more comprehensive financial aid systems for poor college students, but due to the traditional poverty identification methods are subjective, poverty indicators are difficult to quantify and other factors, making the identification of poor students still a difficult problem in college financial aid decision-making. In this paper, based on the campus big data, we analyze the consumption behavior information, living habits and study situation of college students as well as propose a variant of CLS-RFM model. The information entropy modified hierarchical analysis is used to determine the parameter weights from a combination of quantitative and qualitative perspectives on the index factors of students' consumption, life and study, and cluster analysis is performed on students according to the improved RFM variables. In this paper, the model is applied to the actual problem and data to assist a university's financial aid, proving the effectiveness and feasibility of the method.

Keywords: campus big data · poor college students · college financial aid · improved RFM model · clustering

1 Introduction

Under the influence of historical and policy factors such as the integration of colleges and universities, the growing scale of enrollment, and the continuous improvement of the higher education fee system, the number of students from economically disadvantaged families in colleges and universities has been rising rapidly. In response, the government has invested a lot of money every year to ensure that poor college students can successfully complete their studies and enjoy the fruits of higher education development. The scientific and accurate recognition of poor students' qualifications and the realization of accurate financial support for poor students are the major issues that universities must solve in the practice of student support.

X. Yuan et al. (Eds.): ICEKIM 2023, AHCS 13, pp. 1053–1062, 2023.
https://doi.org/10.2991/978-94-6463-172-2_111

At present, the procedure for identifying poor students in most colleges and universities is that the students themselves submit a written application, issue a certificate of poverty from the civil affairs department of the township or township to which they belong, and then the colleges and universities review and determine the list of poor students for financial assistance. However, the traditional identification method is highly subjective, and the authenticity of the submitted materials cannot be guaranteed, and there are many drawbacks such as difficulty in knowing, confirming and proving. With the rapid development of the era of big data, the determination of students' poverty based on the personal feelings and intuition of teachers, counselors and other financial support personnel is often more subjective and the basis of determination lacks persuasive power, while the analysis of a large number of students' school behavior data can make the decision result more scientific and reasonable.

Among the studies exploring the classification of poor students, Zhou et al [1]. Pointed out the problems in the implementation stage of financial aid for poor students in colleges and universities and proposed a logical design for building a support system for poor students based on big data technology. Talingdan [2] used a clustering approach to evaluate poverty data of a community and analyzed families in different subgroups. Sunaryono [3] proposed to combine hierarchical analysis with clustering to deal with the poverty classification problem.

Based on the existing research foundation, this paper introduces customer relationship management to college financial aid work, which uses an improved RFM model, the CLS-RFM model, constructs a CLS segmentation framework for the data information of college students' consumption, living and learning situations. A hierarchical analysis algorithm is also used to determine the weights of each variable in the model considering policy factors, and then clustering algorithm is used to classify students. In this segmentation, the CLS three-dimensional cube example diagram shown in Fig. 1 is constructed to analyze the student profile. In contrast to the use of reviewing written verbal materials, CLS uses big data analysis techniques to identify poor students more convincingly. With the national victory in poverty eradication and the gradual improvement of university financial aid system, the theme of university financial aid research gradually passes from precise financial aid to financial aid for human development, that is, the purpose of financial aid is not only to give material support to poor students, but also to hope that they can be successfully employed and give back to the society. Therefore, CLS not only considers the data of students' consumption behavior in school [4], but also specifies the key sponsored students from multi-dimensional data information mining.

2 Related Work

2.1 Conventional RFM Model

RFM model is a popular analysis model for customer relationship management in marketing, which is mainly used to measure the value of customers and the profit generating ability of customers. R denotes the time interval between the customer's last consumption time and the analyzed time point; F represents the number of transactions made by the customer during a certain period of time; M indicates the total amount of money spent by the customer during a given period of time. In the previous literature [5], an

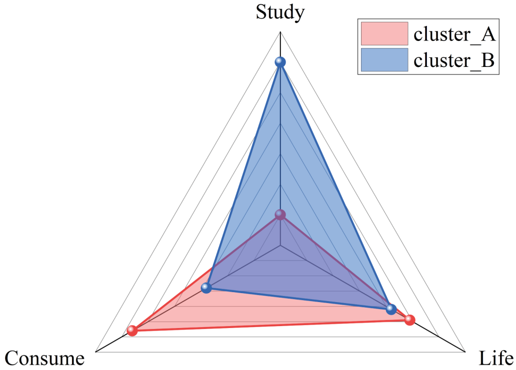

Fig. 1. Example Diagram of Student Segmentation

improved RFM based on consumption data is proposed, which mainly responds to students' poverty in terms of their consumption behaviors such as the number of times and the amount of money spent at school, showing a better characterization to a certain extent.

2.2 AHP Methodology

The hierarchical analysis is a multi-criteria policy-making method proposed by the American operations researcher Professor Saaty [4] in the early 1970s, which mainly transforms ambiguous and qualitative analysis problems into several simple problems for quantitative analysis. In this paper, with reference to the implementation process of hierarchical analysis, the weights of CLS are calculated by hierarchical analysis. However, there is subjectivity in determining the weights of each indicator by hierarchical analysis, and the entropy value method is introduced to correct the determined indicator weights and further objectify the subjective factors.

3 Methods and Processes

3.1 Campus-Based Big Data Processing Student Financial Aid Framework

With the continuous promotion of informatization and digitalization construction in colleges and universities, a large amount of data on students' daily campus life has been accumulated, such as students' canteen consumption, book borrowing, dormitory access control and annual study results. These data can make a more accurate portrayal of college students, thus making the accurate identification of college poor students based on campus big data analysis a reality.

3.2 CLS-RFM Model

The key to financial aid segmentation based on campus big data is the big data analysis method. This paper proposes a new identification model CLS to segment students from various dimensions.

Table 1. Comparison of the meaning of each index of RFM and CLS model

RFM	CLS
R(recency): Last consumption time interval	C(consume): Information on student spending at school
F(frequency): Number of user transactions during the observation time	L(life): Students' attitude towards school life
M(monetary): Total amount spent in a specific time period	S(study): Student learning in school

The definition of poor students at the current stage is not limited to students who show poverty in financial terms, but is often broadened to students with more value for grants, i.e., students with low spending power, positive attitudes toward life, and excellent study performance, among other characteristics. Such students are highly compatible with the national content of precise financial aid, accurate identification, financial support for education and interruption of intergenerational transmission of poverty. In order to improve the accuracy of student identification and classification, this paper proposes an improved RFM model, namely CLS model, based on campus big data. The parameter comparison between RFM and CLS model is shown in Table 1.

The parameters C, L and S of the CLS model are generated by using the data obtained from the data platform such as College One Card, where parameter C indicates the consumption behavior of students in a certain year; parameter L indicates the life attitude of students in a certain year; parameter S indicates the study situation of students in a certain year. In the CLS model, the value of financial support for students is defined as the importance of student clusters in the model indicators, and the higher the value of the corresponding weight value of financial support, the higher the possibility of students accepting financial support. Parameter weight is an important indicator to evaluate the value of student financial aid. In the CLS, a larger parameter weight indicates a greater influence of the parameter on the value of financial aid.

3.3 CLS Analysis

1) Evaluation Indicator System

The personal factors that affect the availability of grants for college students can be focused on the following aspects:

a) The consumption behavior includes the total amount of consumption, the number of times of consumption, the average value of consumption, and the variance of consumption.
b) The attitude towards life includes habitual early rise, habitual late return, number of dormitory punch cards and number of library visits.

c) Study status includes the number of books borrowed, the number of days of effective reading, and academic performance ranking.

Therefore, we define the objective level as the value of student financial support, and the criterion level is divided into consumption, life and study according to the above-mentioned first level indicators. These three categories are composed of 3 or 4 sub-categories of detailed indicators, which we consider as the solution level. The hierarchical model is shown in Table 2.

By the general rule of thumb, the lower the total consumption amount in daily life (C1), the more consumptions in school (C2), the lower the average value of consumption (C3), and the lower the stable output of consumption level (C4), the student is defined as a poor student at the CONSUME level, then C in the CLS assessment index can be expressed as (1), and similarly other index systems are given by (2) and (3).

$$C = -\alpha_1 \times C_1 + \alpha_2 \times C_2 - \alpha_3 \times C_3 - \alpha_4 \times C_4 \tag{1}$$

$$L = -\beta_1 \times C_5 - \beta_2 \times C_6 + \beta_3 \times C_7 + \beta_4 \times C_8 \tag{2}$$

$$S = \chi_1 \times C_9 + \chi_2 \times C_{10} - \chi_3 \times C_{11} \tag{3}$$

where, α, β, χ is the weight of CLS secondary indicators, and it is also known from practice that there are different levels of importance of C, L, S among CLS. In this

Table 2. Financial aid evaluation index system of colleges and universities

Objective layer A	**Criterion layer B**	**Solution layer C**
Value of Student Financial Assistance	Consumption B1	Total annual consumption amount C1
		Number of annual consumption C2
		Consumption mean C3
		Consumption variance C4
	Life B2	Habitual early rising C5
		Habitual late sleeping C6
		Dormitory punch card count C7
		Number of trips to the library C8
	Study B3	Number of books borrowed C9
		Effective reading days C10
		Achievement Rank C11

Table 3. 1–9 Example of Scaling Method Judgment Matrix

Index	C	L	S
C	1	3	1
L	1/3	1	1/3
S	1	3	1

paper, we will use hierarchical analysis to calculate the CLS model parameter weights for each type of student financial aid value.

2) Analysis of indicator weights

a) The relative importance between the indicator variables was measured by using 1 to 9 and their reciprocals as scales. The results of each measure are calculated according to (1), and constitute the judgment matrix.

$$\{M = [a_{ij}]_{n\times n}\}(M = A, B, C) \tag{4}$$

An example of the CLS judgment matrix is shown in Table 3.

b) Consistency test is performed for each judgment matrix, firstly, the consistency index CI is calculated according to (2), and the RI is obtained by finding the consistency index as in Table 4.

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{5}$$

where, $\lambda_{\max}$ is the maximum eigenvalue of the judgment matrix.

If CR < 0.1, the consistency of the judgment matrix $CR = \frac{CI}{RI}$ can be accepted, otherwise the judgment matrix is readjusted. The weights of each index are judged based on the expert ratings, and the judgment matrix is listed using hierarchical analysis. Therefore, the set of weights is derived after the test as W_{AHP}.

$$W_{AHP} = (w_1, w_2, \ldots, w_n) \tag{6}$$

Table 4. Consistency index RI value

n	1	2	3	4	5	6	7	8
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41

c) Modify the set of calculated weights.

To further reduce the subjectivity caused by AHP, m experts are then invited to score the n indicators on a percentage scale, and the evaluation matrix E of m experts for the n indicators can be obtained:

$$E = \begin{bmatrix} e_{1,1} & e_{12} & \cdots & e_{1,j} \\ e_{21} & e_{22} & \cdots & e_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ e_{m,1} & e_{i2} & \cdots & e_{m,n} \end{bmatrix} \tag{7}$$

$$p_{ij} = \frac{e_{ij}}{\sum_{i=1}^{m} e_{ij}} \tag{8}$$

$$H_j = -\frac{1}{\ln m}\sum_{i=1}^{n} p_{ij}\ln p_{ij}(j = 1, 2, ..., m) \tag{9}$$

After normalizing E according to (8) and normalizing it, the information entropy is obtained according to (9) for the indicators.

Then, the entropy weight of each indicator is calculated as:

$$w_{entropy}^{j} = \frac{d_j}{\sum d_j}(j = 1, 2, ..., n) \tag{10}$$

where, $d_j = 1 - H_j$.

Calculate the set of correction weights from (3) and (7).

$$W = \frac{w_{AHP}^{j} + w_{entropy}^{j}}{\sum_{j=1}^{n}(w_{AHP}^{j} + w_{entropy}^{j})} \tag{11}$$

Based on the above judgment of index weights, combined with (1) (2) (3), we can define the value of student financial aid as the sum of the product of each index and its corresponding weight, that is

$$CLS_{value} = \mu_1 \times C + \mu_2 \times L + \mu_3 \times S \tag{12}$$

where, $\mu \in W$.

4 Experimental Data and Analysis

4.1 Data Sources

The data used in this paper comes from a university student database, totaling 6,155 students' one-card consumption data, library checkout data, dormitory access data, library access data and student academic performance ranking data. The CLS specified parameter key field in the data is retrieved from the database table to get the relevant information used for student financial aid breakdown. The data information for one of the students is shown in Table 5.

Table 5. Information column of a student's behavior at school

ID	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
32666	5656.85	1428	3.96	4.87	10	22	78	22	0	0	2836

4.2 Data Standardization

In order to eliminate the influence of different magnitudes of CLS model parameters on student financial aid segmentation results, this paper needs to standardize the collected data.

4.3 Experimental Results and Analysis

The importance of the two-by-two comparison of each indicator was calculated by collecting questionnaires from several experts. Then the 11 indicator system $(C_1, C_2, ..., C_{11})$ was calculated based on hierarchical analysis as W_{AHP} = (0.0556,0.1299,0.0764,0.1667,0.0255,0.0185,0.0433,0.0556,0.0504,0.1963,0.1783). Therefore, the modified parameter weights are W = (0.0767, 0.0888, 0.0667, 0.1669, 0.0603, 0.0581, 0.0692, 0.0791, 0.0741, 0.1221, 0.1381). The CLS model indicator weights are substituted into (1), (2) and (3) to obtain the value of student financial aid as

$$CLS_{value} = 0.3991 \times C + 0.2667 \times L + 0.3343 \times S \tag{13}$$

In this study, based on the Python language, the C, L and S parameters were used as cluster variables, and students were classified into four categories using the K-means clustering algorithm, resulting in the classification results of funded students, as shown in Table 6.

As can be seen from Table 6, the group of students with the greatest value of financial support Level = 1 is group 3, which includes 1574 students, which accounts for 26% of all participants. This group of students has high consumption constraints (the higher the value of the indicator, but the poorer it is). Both life attitude and study situation show a large level, which can be defined as particularly poor and an important group of students for college financial aid, and teachers in the front line of college about financial aid, counselors can focus on this group of students.

Table 6. Results of student segmentation analysis of the CLS model

No.	C	L	S	CLS_value	Level	Number
1	-0.012	0.005	0.051	0.045	2	1894
2	0.004	-0.012	-0.062	-0.070	3	1612
3	0.133	0.011	0.012	0.156	1	1574
4	-0.180	-0.008	-0.016	-0.204	4	1075

The Level = 2 funding group is Group 1 students, which includes 1894 students, accounting for 31% of all students surveyed. This group of students has moderate consumption, but shows excellent life attitudes and learning. This group of students can be defined as relatively poor, a group of students with good character and academic performance, and in line with the national policy point of financial support for education.

The Level = 3 funding group is Group 2, which includes 1612 students, or 26% of the total group. This group of students has excellent consumption and academic performance, belongs to the low consumption group and has excellent academic performance, and can be defined as generally poor.

The Level = 4 category is defined as non-poor students and is rated as weakly funded from the data analysis level, including 1075 students, or 17% of the total. Students in this category are high spenders in terms of consumption, poor life consciousness and weak academic profile. Therefore, this category of students was categorized as average students from the big data level and left for provisional evaluation.

5 Conclusions

Nowadays, the society is in the trend of high-tech development, and it becomes more and more important to use big data analysis to solve problems in life and industry. In this paper, we propose a variant of RFM model to select the daily behavioral data of college students to build a college financial aid framework. By analyzing students' consumption data, library data, access control data and book circulation data, we can comprehensively reflect the importance of students' receiving financial aid. The method proposed in this paper can assist colleges and universities in formulating corresponding financial aid strategies for different value groups of students from the perspective of big data, and avoid the problem of misjudgment and misclassification caused by traditional written applications and oral surveys. It provides certain theoretical support and evaluation basis in practical application.

Acknowledgment. This work was supported by Jiangsu Provincial Education Science Thirteenth Five-Year Plan Project and Project number is X-a/2018/08.

References

1. Zhou, Z., Lu, X., Peng, C., Chen, K., Zhang, F., Wang, Z., ... & Chen, Y. (2020, March). Research on the optimization of the system for the identification, supervision and privacy protection of targeted poverty alleviation for poverty-stricken college students based on big data technology. In 2020 International Conference on Computer Engineering and Application (ICCEA) (pp. 149–153). IEEE.
2. Talingdan, J. A. (2019, February). Data mining using clustering algorithm as tool for poverty analysis. In Proceedings of the 2019 8th International Conference on Software and Computer Applications (pp. 56–59).

3. Sunaryono, D., Akbar, R. J., & Setiyawan, R. D. (2016, October). Poverty classification using analytic hierarchy process and K-means clustering. In 2016 International Conference on Information & Communication Technology and Systems (ICTS) (pp. 266–269). IEEE.
4. Saaty, R. W. (1987). The analytic hierarchy process—what it is and how it is used. Mathematical modelling, 9(3-5), 161-176.
5. Bin, D., Peiji, S., & Dan, Z. (2008, December). Data mining for needy students identify based on improved RFM model: A case study of university. In 2008 International Conference on Information Management, Innovation Management and Industrial Engineering (Vol. 1, pp. 244–247). IEEE.