College students' innovation and entrepreneurship model based on probability theory statistics

Dun Ao, Rigen Mo

1. Inner Mongolia Medical University, College of humanities education, Hohhot Inner Mongolia Autonomous Region, 010001, China
2. Inner Mongolia Medical University, College of Marxism, Hohhot Inner Mongolia Autonomous Region, 010001, China

Abstract

In order to better realize the innovation and entrepreneurship model of college students, the author proposed a statistical method based on probability theory. Based on the data obtained from the questionnaire on "university students in the community of innovation and entrepreneurship", this paper analyzes the existing problems already in the innovation and entrepreneurship of college students in the community. Second, on the basis of pattern analysis and data reliability, the analysis method based on principal analysis is used to determine the eight factors that influence the innovation and entrepreneurship among college students in the community. In addition, innovative and effective business opportunities for community college students based on statistical results have been developed. The test results show that: from the prediction of the last three sets of models, the average relative error of y1 and y2 is 0.0013 and 0.0984 respectively, and the error is less 0.1, indicating that the relative error of the model is within the acceptable range. The results calculated by the model are accurate according to the actual situation from the question. It can be seen that the author uses the analysis method as the principal analysis to determine the eight important factors that affect the innovation and entrepreneurship of the students. High society is necessary, which proves that the design model created by the author is scientific and practical.

Keywords: Probability theory statistics; College students; Business model

AMS 2020 codes: 62H10
1 Introduction

Cultivating college students' innovation and entrepreneurship skills is an important aspect of innovation and entrepreneurship education in colleges and universities. It is not only related to the development of students' work competitiveness and the effectiveness of training college staff, but also related to the use of the country's innovation driven development strategy. In the report sent to the 19th National Congress of the CPC, the state pointed out: "we will create a team of workers with good knowledge, knowledge and innovation, carry forward the spirit of workers and engineers, and create a perfect work relationship and leadership do the best[1]. As the reserve army of the Party and the country in the new era, college students should comprehensively enhance their employment competitiveness. With the continuous development of economy and people's continuous progress, people's knowledge level and professional quality are also constantly improved, at the same time, academic progress, development and improvement also provide a foundation and help for professional talents, in the process of data research and probability analysis, the comprehensive quality of relevant personnel is also constantly improved, meanwhile, in the context of big data, data analysis and statistics often bring great convenience to enterprise development and research. This also provides a variety of sources of research funding, the widening of research funding sources, the research basis will be constantly improved, and the resources of relevant personnel and the research environment will also be constantly improved[2]. In this case, researchers can make better probabilistic analysis and statistics of data, so that the research can be developed more quickly, it provides a guarantee for the accuracy of probabilistic analysis and data statistics, which have certain career prospects under big data, and can bring more material security and living sources for researchers. This also promotes the progress and development of research in a certain sense, and provides a talent base and material guarantee for research.

In the context of "Internet", based on the data obtained from the questionnaire of "university students in the community of innovation and entrepreneurship", the author first analyzes the problems of college students in the community in innovation and entrepreneurship. Second, after analyzing the statistical model and data reliability of the questionnaire data, the analysis method based on principal analysis was carried out to evaluate the principals. which affects the innovation and entrepreneurship of university students in the city. Along with the principal is the process of advising on the statistics of the results, the new models of change and the entrepreneurship of the students in the community based on the statistics of the results. Finally, some suggestions are put forward on how to improve the innovation and the ability of university students in the community [3].

2 Literature Review

According to incomplete statistics from the State Information Center, the success rate of graduates starting a business in 2017 is only 3%-5%, while the success rate of graduates starting a business in European and American countries is as high as 20%-30%. And there's a lot of literature showing that college students have a passion for entrepreneurship, however, due to the influence of their own psychological quality, anti-pressure ability, entrepreneurial skills, management ability and various external factors, the success rate of entrepreneurship of college students is not high. The author took local college students as the research object to conduct a questionnaire survey, in the process of sorting out and analyzing the questionnaire information, the author found the main problems existing in the innovation and entrepreneurship of local college students: Under China's exam-oriented model and "score-only" orientation, the degree of curriculum diversification is not high, homework requirements are standardized, students have less and less practical opportunities, and their desire for innovation and innovation potential are gradually lost, so that college students do not have a clear plan for innovation and entrepreneurship. Figure 1 is about the question "Regarding my future, have
a clear plan for innovation and entrepreneurship ", the answers given by 600 students truly reflect the current situation of local college students' planning for innovation and entrepreneurship.

Figure 1. Is there a clear innovation and entrepreneurship plan

College students' innovation and entrepreneurship ability refers to college students' application of existing knowledge and conditions, working to meet a need or advance society, improve or innovate a thing, method, environment, etc, so that they can be developed and utilized to produce new effects. The cultivation model of college students' innovation and entrepreneurship ability includes three modules: Ability composition, model construction and operation mechanism (see Figure 2), among them, capability composition is the basic content of the model, mode construction is the core plate of the module, and the operation mechanism is an important carrier to realize the module. Wang S discussed two shortcomings of highway traffic data-driven random basic graph. The first disadvantage is related to the least squares method, which is widely used to establish traffic flow basic maps. In our opinion, these methods are not suitable for generating a percentile based random basic graph, because the results generated by the least squares method represent the weighted sample mean, not the percentile. The second disadvantage is the extensive use of independent modeling methods based on percentile basic graph families. The existing methods are not enough to coordinate the basic graphs in the same family, so they do not conform to the basic rules of probability theory and statistics. To solve these problems, this paper proposes a global modeling framework based on the concept of minimizing the mean absolute error[4]. Wu K proposed the method of coefficient of variation in probability theory and statistics to quantify dispersion. Using the coefficient of variation, we calculated the dispersion of the mining blocks and solved the balance problem to quantify the dispersion of Bitcoin and Ethereum systems[5]. Based on this, Hu Q first analyzed the concept and connotation of college students' innovation and entrepreneurship from the perspective of economic transformation, and then proposed the cultivation strategy of college students' innovation and entrepreneurship based on big data analysis[6].
3 Probability Theory and mathematical Statistics analysis of influencing factors of college students' innovation and entrepreneurship

3.1 The ability of probability theory and mathematical statistics to process data is relatively high

Probability theory refers to the analysis and judgment of the possibility of events and the authenticity of events based on existing data, probability is a part of mathematics. There is no doubt about the processing ability of data, and data statistics refers to the storage of data in different categories through relevant operations, and using statistics to better summarize the frequency of various types of events, to a certain extent, probability theory and data processing have some similarities, but data statistics is a form of data processing formed on the basis of modern technology, both parties classify and analyze data by certain means, in the process of data application, probability theory prefers the possibility of events, data statistics tend to count the sum of events, and the two can be used in a complementary manner, by cooperating to play a greater effect, probability theory is based on data statistics for analysis, through data statistics to make probability more accurate, data statistics are the support of probability theory and provide data basis and simple processing for probability, as a result, data are analyzed and classified through data processing, which makes probability analysis more concise, convenient, efficient and high-quality[7].

3.2 Integration of data statistics and probability

The continuous application of the network makes big data appear information explosion, and more and more data appear on the network. Although these data can bring people a lot of help in life and work, but because of the huge amount of data, people have difficulty in selecting and screening, and more data means that people need to spend more time in selecting the information they need. At the same time, the same individual will make different choices due to different positioning in different social environments, when personal identities are constantly changing, it becomes more difficult to process and filter data[8]. In order to better ensure the efficiency of data use, statistics of the data in today's society, probability and data statistics have become necessary projects, only through probability and mathematical statistics can social phenomena and people's general psychological
needs be better reflected, at the same time, through probability and data statistics, people can better improve the efficiency in the process of applying big data[9].

3.3 Data Sources

The author adopted the form of online questionnaire, 600 valid questionnaires were collected from 3 985, 211 and 9 ordinary local colleges and universities nationwide. The subjects of the survey involved college students of literature and history, science and technology, art, sports and other colleges, including 309 female students and 227 male students.

3.4 Statistical analysis of samples

In order to further analyze the influencing factors of local college students' innovation and entrepreneurship, the author selected school type, major type, college students' understanding of innovation and entrepreneurship, capital problems, experience and social relations, whether there is a clear plan for innovation and entrepreneurship, whether the family can provide a platform, how much the school pays attention to the cultivation of students' awareness of innovation and entrepreneurship, the school offers systematic education courses on innovation and entrepreneurship theory, the practice training of integrating innovation and entrepreneurship into the classroom, whether there are full-time innovation and entrepreneurship instructors, the participation of college students in innovation and entrepreneurship activities, the participation of people around them, local government's support for college students' innovation and entrepreneurship, relevant policies and documents, the 16 questions about the participation of other units in the school's innovation and entrepreneurship activities were designed and investigated through the questionnaire[10]. All the above questions in the questionnaire were in the form of Richter scale, with 5 points set to measure the positive or negative response of a statement, that is, each option was given a score, the scores from very good to very bad are 5, 4, 3, 2, and 1. In general, Richter scales have higher reliability than scales of the same length. For the convenience of narration, let each question correspond to a variable, and the 16 variables are denoted as X1, X2,... X16. The frequency analysis method was used to analyze the general characteristics of the respondents from four aspects of gender, major type, school type and grade, and the basic information was obtained in Table 1.

| Table 1. Basic information of the survey respondents |
|----------------------------------------|----------------|----------------|----------------------------------------|
| Statistical variables                | Taxonomic category | Valid percentage of % | Statistical variables | Taxonomic category | Valid percentage of % |
| Sexual distinction                   | Female          | 42.4            | School type               | 985/211          | 27.7               |
|                                      | Male            | 57.6            | Ordinary undergraduate   | 24.4             |                   |
|                                      |                 |                 | Second level undergraduate| 29.8             |                   |
|                                      |                 |                 | Grade III undergraduate and junior college| 18.1 |                   |
| Professional type                    | Humanities and Social Science | 26.6          | Grade                    | Freshman         | 28.4               |
|                                      | Natural science | 26.5            | Sophomore                | 12.1             |                   |
|                                      | Engineering and Technology | 32.3          | Junior                   | 25.5             |                   |
|                                      | Eurhythmics     | 14.6            | Senior                   | 33.9             |                   |
Statistics in Table 1 show that: The distribution of these four items of information is relatively balanced, and there is no large deviation, which indicates that the data obtained by the questionnaire is valid.

3.5 Reliability analysis of sample data

According to the reliability and validity test, the consistency, stability and reliability of the questionnaire data obtained by the author can be detected, which can avoid the big deviation of the final results[11]. The author chose the commonly used Cronbach's $\alpha$ coefficient method for the intrinsic reliability test, the higher the value of the reliability coefficient $\alpha$, the higher the reliability of the questionnaire data. In general, the classification of reliability levels is shown in Table 2.

<table>
<thead>
<tr>
<th>Grade</th>
<th>limits</th>
<th>0. 60—0. 65</th>
<th>0. 65—0. 70</th>
<th>0. 70—0. 80</th>
<th>0. 80—0. 90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unlikelihood</td>
<td>Minimum acceptable value</td>
<td>Fairly good</td>
<td>beyond compare</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cronbach's $\alpha$ reliability coefficient was calculated by the following formula:

$$\alpha = \frac{p}{p-1} \left( 1 - \frac{\sum_{i=1}^{p} \sigma_{p}^{2}}{\sigma_{x}^{2}} \right)$$

(1)

Where $\alpha$ is the reliability coefficient; $p$ is the total number of assessment questions set in the scale; $\sigma_{p}^{2}$ is the in-table variance of statistical results corresponding to the LTH evaluation question; $\sigma_{x}^{2}$ is the variance of the sum of the statistical results of all the evaluation questions. According to Equation (1), Table 3 shows Cronbach's $\alpha$ coefficient of influencing factor data of local college students' innovation and entrepreneurship, the reliability coefficient based on the standardized term is 0.745, which is between 0.70 and 0.80, therefore, it can be seen that the internal reliability of the recovered data of the questionnaire is relatively ideal[12].

<table>
<thead>
<tr>
<th>Cronbach's $\alpha$</th>
<th>The Cronbach's $\alpha$ based on the normalization term</th>
<th>Number of terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. 748</td>
<td>0. 745</td>
<td>20</td>
</tr>
</tbody>
</table>

3.6 Establishment of comprehensive model of local universities based on probability statistics

According to the relationship between input layer and output layer of probability statistics, a comprehensive model of the problem is established as follows.

Step 1: Establish the artificial neuron. In the training process of adaptive reinforcement learning, different activation functions were selected successively, the learning rate and weight were constantly adjusted, and the tangent sigmoid tansig function was finally selected as the activation function:

$$\tan \operatorname{sig}(z) = f(z) = \frac{1}{1 + e^{-2z/k}}$$

(2)
Where z represents each principal component variable; β is the parameter of sigmoid function.

Step 2: In order to calculate the correlation degree between the input layer (8 principal components extracted by factor analysis method) and the output layer (recognition of local college students to the university and government), the calculation formula of $R_{mn}$ is defined as

$$R_{mn} = \sum_{s=1}^{14} W_{K,m} \frac{1-e^{-\gamma_{nK_s}}}{1+e^{-\gamma_{nK_s}}}$$  \hspace{1cm} (3)

Where $K_s (s = 1,2,\ldots,14)$ is the hidden layer neuron, $K_m$ is the connection weight between the neuron corresponding to the input component $z_m (m = 1,2,\ldots,8)$ and the hidden layer neuron $K_s (s = 1,2,\ldots,14)$, $\gamma_{nK_s}$ is the connection weight of hidden layer neuron $K_s (s = 1,2,\ldots,14)$ and input component $y_n$ [13].

Step 3: Calculate the influence between the input layer and the output layer.

$$F_{mn} = \frac{R_{mn}}{\sum_{m=1}^{8} \sum_{n=1}^{2} R_{mn}}$$  \hspace{1cm} (4)

According to steps 2 and 3, the comprehensive model of innovation and entrepreneurship of local college students based on probability theory statistics is as follows:

$$F_{mn} = \frac{\sum_{s=1}^{14} W_{K,m} \frac{1-e^{-\gamma_{nK_s}}}{1+e^{-\gamma_{nK_s}}}}{\sum_{m=1}^{8} \sum_{n=1}^{2} \sum_{s=1}^{14} W_{K,m} \frac{1-e^{-\gamma_{nK_s}}}{1+e^{-\gamma_{nK_s}}}}$$  \hspace{1cm} (5)

4 Simulation experiment and analysis

The author uses MATLAB7.0 software as a programming tool, on the experimental platform with 1.70GHz Windows processor, Intel(R)Penium(R)3558UCPU and 8.0GB memory, the 8 principal components $z_m (m= 1,2,\ldots 8)$ as the input layer of the comprehensive model, the recognition degree of local college students to the university and government $y_n (n= 1,2)$ is taken as the output layer of the comprehensive model, the validity of the model is verified by simulation[14]. After calculation, at the 76th iteration, the variance of simulated data is 0.000907, which meets the requirements of learning accuracy, and the variance of training results is output, as shown in Figure 3.
Figure 3. simulates the variance

Through the prediction of the last three groups of samples, the average relative error of \( y_1 \) and \( y_2 \) is 0.0013 and 0.0984, respectively, and the error values are all less than 0.1, indicating that the relative error of this model is within the acceptable range.

Put 600 sample data into Equations (3) and (4), and the correlation degree and influence degree can be calculated as follows:

\[
R_m = \begin{cases} 
R_1 = 6.4831 \\
R_3 = 3.9106 \\
R_2 = 3.7070 \\
R_4 = 2.5367 \\
R_6 = 2.2879 \\
R_5 = 1.0318 \\
R_8 = 0.9408 \\
R_7 = 0.8979 \\
\end{cases}, \quad F_m = \begin{cases} 
F_1 = 32.19\% \\
F_3 = 21.00\% \\
F_2 = 18.59\% \\
F_4 = 12.59\% \\
F_6 = 10.22\% \\
F_5 = 4.89\% \\
F_8 = 0.31\% \\
F_7 = 0.21\% \\
\end{cases}
\]

It can be seen from the calculation results that: The comprehensive results calculated by the model are basically consistent with the actual situation reflected by the questionnaire. This shows that the eight main factors that affect the innovation and entrepreneurship of local college students determined by the factor analysis based on principal component analysis are reasonable, and also shows that the comprehensive model established by the author is scientific and practical[15].

Through data processing and probability analysis, the content in the data can be better refined, so that the professional research has a wider range of forensic space, for the depth and breadth of professional research to provide conditions. In the future of big data, people can also take into account more aspects of problem analysis, which can not only analyze the characteristics and characteristics of an era, at the same time, through data processing and probability analysis of big data, we can analyze the characteristics and conditions of individuals in the current era, for every man is subject to the society of which he belongs and is at the same time a single individual, with different research
directions, forensics on the nature of people are often different, and the basis of big data can give consideration to both the whole and the details[16]. It can not only analyze the universal characteristics and phenomena of an era, but also deeply explore the characteristics and characteristics of individuals, and achieve the ability to take into account the commonality and personality. In academic research, the specialized analysis of specialized disciplines can also be achieved, at the same time, the relevant research can be broader, and all aspects can be taken into account to expand from the major as the main line, so as to supplement the fields involved[17-19].

5 Conclusions

In order to better ensure the accuracy and effectiveness of data, it is necessary to carry out probability analysis and data statistics for big data, at the same time, the continuous crossing of disciplines will also provide the production environment provided by big data, in the environment of big data, the demand for research and analysis of probability analysis and data statistics is increasingly high, on the one hand, probability analysis and data statistics can better feedback the reality and make information acquisition more quality and efficient. On the other hand, big data also needs the integration of data statistics and probability analysis means to make people more convenient, concise and effective in data acquisition. Based on principal component factor analysis, the author discusses the main influencing factors of local college students’ innovation and entrepreneurship, and finds that individual effort is the main aspect affecting college students’ innovation and entrepreneurship. Combined with the existing problems in the innovation and entrepreneurship of local college students, the following gives some suggestions on how to effectively promote the innovation and entrepreneurship of local college students: Theory is the forerunner of practice, as a new era of college students, in the process of learning should first consolidate theoretical knowledge, constantly strengthen learning, enrich themselves; Besides study, I should also actively participate in the innovation and entrepreneurship activities organized by the school, adhere to the application of theoretical knowledge to practice, enrich my practical experience, and constantly improve my innovation and entrepreneurship ability and practical ability. The process of starting a business may be difficult and long, college students should have persistent pragmatic spirit and indomitable spirit to start their own business.

References


