

A Statistical Analysis of the Social Status of Chinese Women: Data Mining Method Based on MapReduce

Liu Xiangdong¹, Yang Mingzhu and Luo Chengshuai

Department of Statistics, Jinan University, Guangzhou, China

Abstract. By studying the influencing indicators of women's social status, we perform an ordered Logit regression analysis on the data of the China Comprehensive Social Survey in 2012, 2013 and 2015, and then select the assessment of self-social status in the female sample as the dependent variable. Using the impact indicators as independent variables to explore the impact of each variable on women's social status. At the same time, applying k-means clustering analysis based on MapReduce to mine the relationship between employment and education level between different genders. We find out the fact that women have a high level of education does not necessarily result in good employment treatment. Gender discrimination in the Chinese labor market is also persistent.

Keywords. Women's social status; ordered logit regression; MapReduce model; k-means cluster analysis

1. Introduction

In recent years, China and society have carried out a series of activities to protect women's rights and promote gender equality. The "12th Five Year Plan" of national population development takes promoting gender equality, family harmony and advocating freedom and equality of marriage as the main task, and improves laws and regulations to establish a solid judicial guarantee for gender equality and gender rights and interests.

In February 2019, nine departments, including the Ministry of human resources and social security and the Education Bureau, issued the notice on "further standardizing recruitment and promoting women's employment", which stipulates that discriminatory behaviors such as "asking about women's marriage and childbirth", "limiting gender or gender priority" and "restricting childbearing as employment conditions" are prohibited in the recruitment process, and also expresses the need to strengthen the human resources market supervision, establishment of joint interview mechanism and improvement of judicial relief mechanism to ensure the "zero existence" of gender discrimination in the recruitment process.

But a new report on women's employment published by the United Nations International Labor Organization in 2019 shows that it will take more than 200 years to achieve gender equality in domestic work. The report points out that although the

¹ Liu Xiangdong, Department of Statistics, Jinan University, Guangzhou, China; E-mail: tliuxd@jnu.edu.cn

proportion of women in management is increasing, it is still low. According to the 2018 report of the world economic forum, the overall gender gap has narrowed by 3.6% since 2006, but it will take 108 years to achieve gender equality globally. In 2018, China ranked 103rd in the global gender equality ranking.

It can be seen that although the social status of women is gradually improving with the improvement of the overall awareness of gender equality, the current situation is still not blindly optimistic. While encouraging and supporting the existing policies, it is still necessary to examine the current situation of gender equality through empirical research. By optimizing and refining the indicators for measuring women's social status, improving the methods of horizontal and vertical comparison, we can draw more practical conclusions and make suggestions for the better development of society.

In the study of women's social status, we can find that the issue of employment equality between men and women is crucial. In the research results, researchers have generally reached a consensus that gender discrimination is an important factor causing gender differences. Material resources and social resources reflect the actual status, so we should take the degree of material resources and social resources actually controlled by women as a measure of their social status (Dixon, 1978) [1]. Women's social status is a concept determined by a reference (Chen, 1993) [2]. Therefore, in most of the studies on women's social status, researchers usually understand women's social status from the perspective of gender structure, and take men's social status as the comparative object. We should compare with men from all aspects of society, politics, economy, culture and family life, taking such equality as the measurement standard to measure women's status (Wei, 1999) [3]. We should classify women's social status by the degree of being respected by others and whether they have equal opportunities for survival and development (Ye, 2003; Liu, 2003) [4].

MapReduce programming model has advantages in big data processing, which can greatly shorten the running time, make programming easy to use. It only needs to run on a large number of cheap computer clusters, and the requirements for hardware devices are not high (Liu, 2011) [5]. In the study of graduates' employment, previous researchers found that random forest is the best for the classification accuracy of graduates' employment destination by classifying various training set data. Based on the decision-making system framework of Guruler in 2010 (Guruler, 2010; Istanbulu, 2010) [6], the employment information recommendation department (Tao, 2012, Leting, 2012) [7], and the incremental overall learning model of ADE based on majority voting decision in 2014 (Ade, 2014) [8], Qiao Fei and others further proposed that the data in the sample set of students' employment training are basically of label type, and the distribution of attribute values was quite different, so the classical random forest model was no longer applicable, they improved the classification performance of the student employment target data set, and combined the current big data background, proposed a distributed improvement model based on MapReduce (Parmar, 2015; Qiao, 2017) [9,10]. Sun Xiaoyu and others used ordered logit model and marginal effect analysis to study the influence and degree of each factor on women's social status. They also focused on the analysis of the connotation of women's social status, and discussed the indicators to measure women's social status. They further believed that the academic community ignored women's self-recognition and self-evaluation, so they studied the factors affecting women's social status from the social, family and personal levels in the previous research framework (Sun, 2018; Zheng, 2018) [11]. William interviewed employees in the workplace, used OLS regression method to study gender discrimination in the workplace, and hoped to use this method to examine the

effectiveness of various policies in the workplace. They also stressed the seriousness of the idea of gender discrimination at work and pointed out the possibility of tracking and investigating the change of this idea (William ,2019) [12].

Based on the above characteristics, we will carry out the distributed data mining based on MapReduce, classify the employment data of graduates and explore the relationship between the education level and employment destination between different genders.

2 model and method

2.1. logit model

We regard y_i^* as the real social status of the i th interviewee, and as an unobservable potential variable. The ordered logit model can map it to the ordered variable y , and the relationship between the two can be expressed as follows:

$$y_i = \begin{cases} 1, \gamma_0 < y_i^* \leq \gamma_1 \\ 2, \gamma_1 < y_i^* \leq \gamma_2 \\ \vdots \\ j, \gamma_{j-1} < y_i^* \leq \gamma_j \end{cases} \quad (1)$$

Where $\tau = (\gamma_0, \gamma_1, \dots, \gamma_j, \dots, \gamma_j)$ is the set of social status classification points of women, all of which are unknown. At the same time, various factors affecting y are expressed in X , and the expression of the model is constructed as follows:

$$y_i^* = \alpha + BX_i + \varepsilon_i \quad (2)$$

Among $X_i = (x_{i1}, x_{i2}, \dots, x_{ik}, \dots, x_{iK}; i = 1, \dots, N; k = 1, \dots, K)$ are the vectors of the observed values of the influencing factors on women's social status in each valid sample. N is the number of valid samples and K is the number of selected key variables. $B = (\beta_1, \beta_2, \dots, \beta_k, \dots, \beta_K)$ is the influence degree of each variable on the dependent variable, and is also the object which I will explore. ε_i represents the random error term, including the sum of all factors not included in the model or difficult to observe.

2.2 main ideas of MapReduce

Based on MapReduce, there are many algorithms that can be applied, such as clustering, decision tree, naive Bayes and so on. As a unsupervised machine learning method, clustering does not need to use training set. Therefore, based on the purpose of exploring the relationship between gender and employment, combined with the consideration of practicability and operational feasibility, this paper selects K-means clustering algorithm as the parallel algorithm based on MapReduce.

1) Data segmentation stage

According to the size of the data and the computing power of each data node, the clustering task data block is reasonably divided, and these maptasks are allocated to each node.

2) Mapper stage

In this part, the map function is used to decompose the task.

The input data blocks are executed on these nodes by the map function written by the operator, that is, the input key / value pairs are processed, and a new set of key / value pairs is generated. Such new key / value pairs are symmetrical as intermediate results, which are not output and written to the local hard disk.

3) Reducer stage

This part uses the reduce function to merge tasks.

Through network transmission or remote program, the key / value pair from map operation is input to the node where the reduce task is running. Here, the reduce function written by the operator will merge the data sets with the same "key" and store the result output on HDFS. HDFS refers to the distributed file system, which can manage cross computer network storage and is also a sub project of Hadoop.

4) Combiner stage

This part uses the combine function, running between the map function and the reduce function.

Processing key / value pairs from map operations can reduce the intermediate results of map output, thus reducing the network transmission load. The form of combine function is generally the same as that of reduce function. The difference is that the output key / value pair type of combine function should be the key / value pair type of intermediate result, so it can be processed by reduce function.

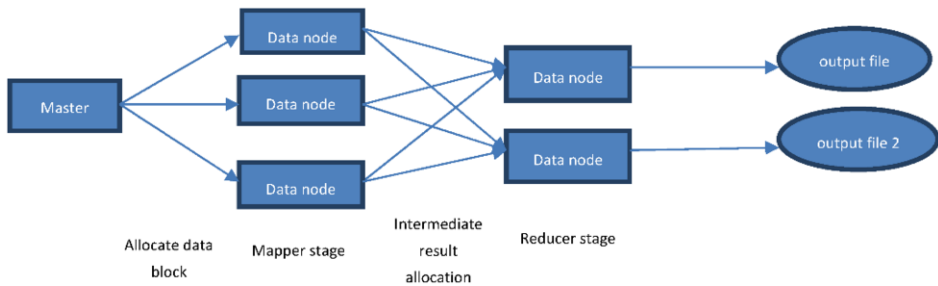


Figure 1. MapReduce calculation framework.

2.2.1 improved k-means clustering model

1) Traditional K-means algorithm

The process of K-means algorithm is to determine the initial center point according to the cluster, calculate the distance between each data point and each center, so as to preliminarily divide the cluster. For each cluster, the average value is re-taken as a new series of centers, and then the distance between each data point and all new centers is calculated, so as to re-divide the cluster, and repeat the process of re-taking the center point to divide the cluster until each cluster no longer changes. The algorithm inputs the database with n data objects, the number of k clusters, and outputs k clusters with the least square error criterion.

2) Canopy clustering algorithm

In the traditional K-means algorithm, it is easy to find that the calculation efficiency is related to the number of processed data, the number of clusters and the number of iterations required. Therefore, in order to improve the efficiency of clustering, the number of iterations must be reduced. On the basis of parallelization, in the face of massive data, canopy algorithm can be used to better determine the value of

K. at the same time, the inaccurate measurement method is used to divide the data points into overlapping areas, so as to generate canopies set and reduce the total amount of calculation.

The canopy clustering algorithm is divided into two stages. In the first stage, the data is divided into overlapping subsets, namely the canopies set mentioned above. In the second stage, we need to use accurate standard clustering method to calculate the distance of canopy as the midpoint. Therefore, canopy algorithm only needs accurate calculation in the second stage, which can reduce the amount of calculation, because the sub datasets divided in the first stage can overlap, so it also increases the fault tolerance of the model.

The process of the algorithm is as follows: first, put all data points into the list m , and select a point n in the list. For each canopy c , if the distance between point n and c is less than the preset threshold T_1 , it will be classified as canopy c . If the distance is less than the preset threshold T_2 ($T_1 > T_2$), then the point is a strong marker. If there is no strong mark in the process, it need to generate a new canopy. Repeat for all data points. Each canopy will generate a list of center points. Canopy clustering algorithm input point n , preset threshold T_1, T_2 ($T_1 > T_2$), output the center point of clustering.

K-means algorithm is based on canopy algorithm. In the first stage, regarding each canopy as the center of the initial clustering, and in the process of repeatedly calculating the distance between data points and the center, only need to calculate the distance between data points and the center in the same canopy.

2.2.2 parallel design of computer cluster

1. Parallel design of canopy algorithm

(1) Design of map stage

This stage focuses on the setting of canopy threshold. Compress the distance formula of strong correlation into the square difference absolute value open root sign, enlarge the distance formula of weak correlation into the square of the square difference absolute value, and calculate the average distance between all sample classifications according to such distance formula. According to the above classification number, it need a total of $(C130)^2$ operations, and take the average value as the average distance AD . According to the theory and experience, AD is much higher than the threshold we need, so we need to use the threshold coefficient to compress. Such coefficient is related to k value, so we use previous research to make the assumption that the sample distance in sample space is uniform, set the threshold coefficient to $k/2$, and set the calculation of threshold to $\frac{AD}{2}$, and modify the threshold through testing.

(2) Reduce stage design

Collect the center points of the output in the Map stage, then merge them according to the key, and the output set of center points is the initial center point in the K-means clustering stage.

2. Parallel design of K-means algorithm

(1) design of map stage

For all the data in the data set, the distance between each data point and the initial clustering center output by canopy algorithm, then calculate and allocate them according to the shortest distance. And the process will continue to iterate until it can no longer be carried out.

(2) design of reduce stage

Combine the intermediate results of map stage with reduce function, and output cluster center and cluster results. A summary of the model framework is shown in Figure 2.

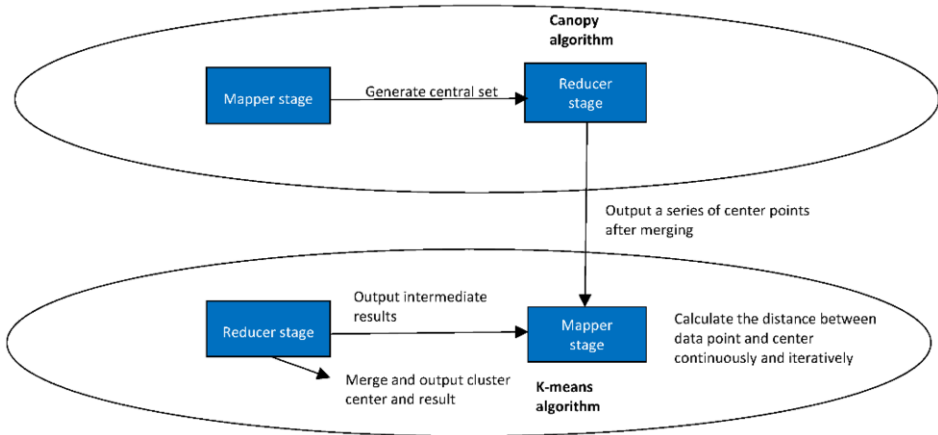


Figure 2. Model framework summary.

3 empirical analysis

3.1 empirical analysis of Logit Model

3.1.1 data source and background

We use the data of China Comprehensive Social Survey (CGSS) in 2012, 2013 and 2015 to analyze women's social status and its related influencing factors, including horizontal analysis and vertical analysis. According to international standards, China's comprehensive social survey has conducted a continuous cross-sectional survey of more than 10000 households in various provinces and autonomous regions of the mainland of China every year since 2003. In principle, these surveys use multi-level stratified random sampling. In the second phase of CGSS, it selects a total of 100 county-level units plus five metropolises, 480 villages / neighborhood committees and 12000 individuals in China, and carries out a follow-up survey. The survey implementation was representative and the data quality was high.

3.1.2 variable description and assumption

In the horizontal analysis, we aim at the evaluation of women's social status and the research purpose of its influencing factors. On the basis of predecessors, it takes the evaluation of women's social status as the dependent variable, and the dependent variable "women's social status" in the questionnaire takes an integer of 1-10 (including 10) as the score, with the score of 1 as the lowest level and the score of 10 as the highest level. The score is divided into three grades: 1-3 (including 3) as the "low" grade; 4-7 (including 7) as the "medium" grade; 8-10 (including 10) as the "high" grade.

At the same time, we take family, society and individual as the three aspects of dependent variable research. Among them, family aspect includes family education background, family burden, husband and wife difference; social aspect includes personal political participation degree, household registration nature and current employment situation; personal aspect includes human capital, self-recognition and gender awareness.

Therefore, in the questionnaire, the education level of the mother, the education level of the father, the number of children born and the number of the elderly in the family, the education level of the husband and the wife, the income difference between the husband and the wife, whether they participate in the election, whether they are non-agricultural registered permanent residence, whether they are employed, the age, the degree of personal education, the total annual income of the individual, whether they are happy, and the social status changes compared with many years ago. The key variables of the study are perception of future social status, self-confidence, gender division of labor, gender advantage, self-concept, gender discrimination and household division of labor.

3.1.3 Data processing and descriptive analysis

Deal with missing values first. After extracting the key variables, because the data volume itself is large but does not have the conditions to give the simulation value, no matter the mean value replacement or multiple replacement of the missing value, it is easy to cause the error and bias of the research results, so I deleted the data with missing value of the key variables in the processing.

Then select the sample. As the social survey is based on households, the survey will ask the interviewees about the subjectivity of the key variables, so the gender of the interviewees is random, then overall is statistically significant. I study the relationship between women's status and various possible influencing factors, so when processing the data, we exclude the male sample and take the subjective status evaluation of the female sample as the dependent variable of the study.

3.1.4 model results

In the process of model analysis, treating the value of income with logarithm, and expressing the degree of income difference between husband and wife by logarithm difference. Using Stata to analyze the data, the results are shown in table 1.

Table 1. Estimation results of ordered logit model for influencing factors of female social status

Variable name	2012		2013		2015		
	coefficient	Standard error	coefficient	Standard error	coefficient	Standard error	
Family education background	Mother's education	0.002	0.027	0.030**	0.0124	0.024*	0.0142
	Father's education	0.060***	0.021	0.010	0.0113	0.0137	0.0125
Family burden	Number of children	0.089**	0.035	-0.002	0.0362	0.044	0.029
	Number of the elderly in the family	0.187***	0.050	0.050	0.050	0.051	0.053

Differences between husband and wife	Difference in educational level between husband and wife	0.064***	0.013	0.077***	0.013	0.055***	0.014
	Income difference between husband and wife	0.103***	0.020	0.070***	0.0142	0.056***	0.012
Political participation		0.267***	0.069	-0.057	0.071	0.151**	0.076
Account character		-0.035	0.035	0.008	0.064	0.160***	0.060
Employment		-0.014	0.083	-0.253***	0.087	-0.127	0.089
Personal capital	Age	-0.003	0.004	0.009**	0.004	0.007*	0.004
	My education level	0.097***	0.012	0.090***	0.013	0.084***	0.013
	My total income for the year	0.111***	0.021	0.092***	0.015	0.065***	0.019
Self-recognition	Happiness perception	0.584***	0.052	0.599***	0.054	0.733***	0.059
	Vertical change of social status	0.757***	0.050	0.751***	0.051	0.774***	0.054
	Self confidence in social status change	-0.135**	0.061	-0.158**	0.063	-0.003	0.066
Gender consciousness	Gender division of labor	-0.002	0.043	0.080*	0.046	0.041	0.049
	Gender advantage	0.039	0.041	-0.082*	0.044	0.022	0.048
	Sense of autonomy	-0.039	0.040	-0.069	0.043	-0.098**	0.046
	Sexualgender discrimination	0.040	0.053	-0.010	0.054	0.113**	0.057
	Division of housework	-0.035	0.050	0.153***	0.050	0.029	0.056
/cut1		1.979	0.289	1.694	0.263	1.736	0.257
/cut2		6.728	0.314	6.556	0.290	7.143	0.295

Note: * means significant at 10% significance level, ** means significant at 5% significance level, and *** means significant at 1% significance level.

3.1.5 Overall result analysis

Comparing the 3-year coefficients with the standard errors one by one, and comparing the significance, we can find that the estimated results are basically stable. Under the guarantee of this robustness, I can analyze the following three aspects from family, society and individual.

1) family

We can find that in 2012, family variables have a significant impact on women's social status at a certain level, but in 2013 and 2015, the impact of family burden on women's social status is no longer significant, which is to a certain extent the performance of women's gradually decentralizing the focus of life from family to individual or society. Further analysis shows that there is a positive correlation between parents' education level and women's social status. However, this relationship has been

weakened from 2012 to 2015, which can be seen as the result of the popularization of the Internet, and education no longer relies on family education.

There is also a positive correlation between family burden and dependent variables. We interpret this phenomenon as that the heavier the family burden, the harder the individual in the family needs to work and have more opportunities to participate in social activities, which brings a sense of achievement for women, thus improving their self-evaluation of social status. There is a negative correlation between the number of children born in 2013 and dependent variables, which may be related to gender self-awareness and subjective choice in social reality. Women with higher social status tend to choose fewer and better children.

2) social aspects

The relationship between the degree of political participation and the dependent variables was significantly positive at the level of 1% in 2012 and 5% in 2015, but not significantly negative in 2013. On the whole, the higher the degree of political participation, the more able to voice for their own rights and interests, so as to further promote the improvement of social status, get a better platform for political participation, have more opportunities for political participation, is a process of positive feedback.

In terms of employment, whether women are employed has a negative correlation with their social status, that is, the more unemployed women are, the higher their social status is. This may be because women with stronger self-awareness are more inclined to study and further study, so that they can understand gender equality and clarify their social status. However, in the current human resource market, there may still be sexism. Even if women find jobs, they may not be able to really bring them a sense of self-identity, or their wages are lower than that of men in the same job.

3) Personal aspect

There is a strong positive correlation between education and income and women's social status. Education can help people to set up values, and wealth is the most easily noticed social capital. The more wealth and income, the more able to meet their own desires.

In 2013, the gender division of labor had a significant impact on women's social status at the level of 10%, but in 2015, the impact was no longer significant. It may be because the role orientation of men and women in marriage is no longer a fixed model, but a reasonable distribution according to their strengths. In 2013, gender advantage also showed a significant impact on women's social status. The more women disagree that "men are born stronger than women", the higher their social status. In terms of the concept of self-determination, the more women can pay attention to their own development rather than "whether they marry well", the stronger their self-awareness and social status are. Similarly, the higher the social status of women who can think that housework should be shared equally, and data from 2013 confirm this conclusion.

However, in terms of gender discrimination, the data in 2015 showed that it had a significant impact on women's social status, and the social status of women who had gender discrimination in the workplace was also high. We can explain this by the fact that in the sample of women, many high-level women are still affected by the patriarchal society, and they are still inclined to dismiss women in the economic downturn. This also shows that in the process of eliminating gender discrimination and creating human rights equality, we should not only overcome the prejudice of another group, but also pay attention to the prejudice against compatriots within the group. This requires ideological guidance and education.

On the whole, family is no longer the main factor that affects women's independent development and seeks equal rights, while society and individual have increasingly significant influence on women's social status.

3.2 Gender and employment data mining based on MapReduce

3.2.1 purpose and significance

After exploring the influencing factors of women's social status, it is necessary to analyze the role of gender equality initiatives and policies in all aspects of society in recent years. As one of the four goals of China's macro-economic policy, full employment has always been the focus of research. Whether to achieve equality and non-discrimination in employment is an important measure of gender equality and one of the embodiment of women's social status.

Therefore, combining with the background of big data, using the data mining algorithm based on MapReduce to mine the relationship between employment and education level between different genders and classify the data can not only expand the ideas for the use of this method, but also help us understand the basic situation of employment direction between different genders in the past few years.

3.2.2 data processing

We use the survey data of China Comprehensive Social Survey (CGSS) in 2015, and select gender, education level, employment, nature of work and nature of work unit as the research objects and key variables.

First, merge the information. After merging the information of interviewees with the employment information of their spouses, there are a total of 19749 gender and employment data. Then delete the data. Deleting the samples with missing values of key variables, and screening out the graduates of compulsory education. The number of valid samples after deletion is 7285. Reset and quantify the data. For example, the sample of "labor for the purpose of obtaining economic income" but lack of variables such as the nature of work is identified as "farming". Classifying the working nature, type of working unit and ownership of the sample in a standard way. For gender, men and women are represented by 0 and 1 respectively.

All valid samples were classified by gender. There were 2826 valid samples for women and 4459 valid samples for men. The main reason why the number of effective samples is not equal is that the sample without work is directly deleted. Obviously, the sample of women is more than that of men.

3.2.3 empirical results

In the process of demonstration, using the open-source platform of Hadoop to build the overall environment. The output results are shown in the following figure 3 and figure 4.

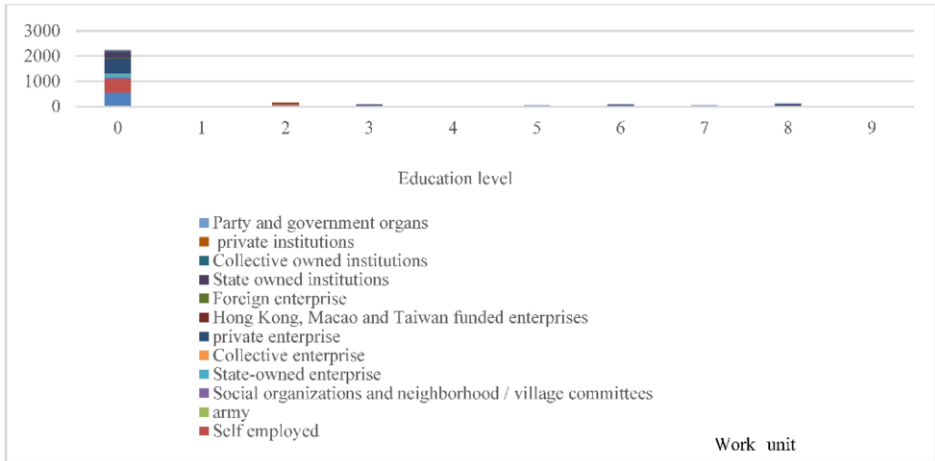


Figure 3. Cluster analysis results of female samples

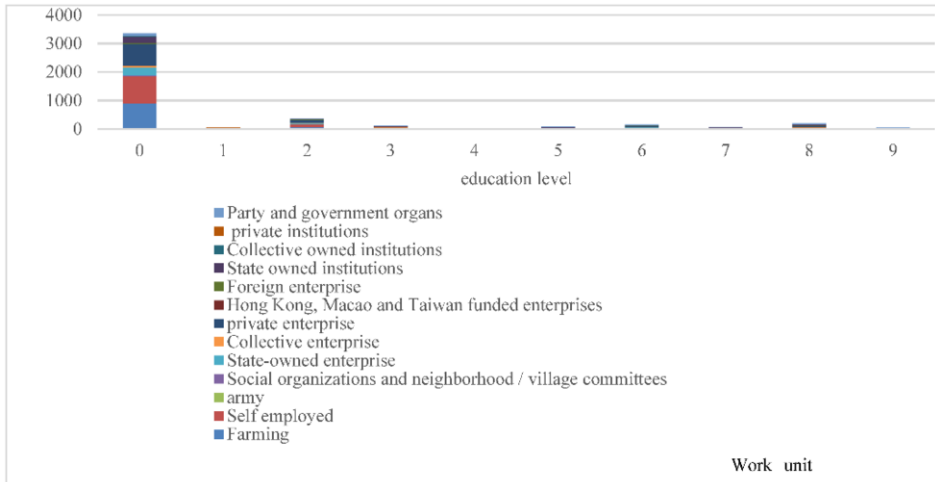


Figure 4. Cluster analysis results of male samples

From Figure 3 and Figure 4, we can see that the number of people employed after the completion of compulsory education accounts for the vast majority of the total number of samples, and this part of the population is widely distributed in all work units, and Agriculture, self-employed and collective enterprises are the most widely distributed. We think it is because some companies will still recruit low-educated people for less knowledgeable jobs. In addition, the number of high school graduates and undergraduate (formal higher education) graduates reached the peak. Therefore, the samples of Higher Education (that is education level = 5, 6, 7, 8) will be analyzed separately. As shown in Figure 5 and Figure 6.

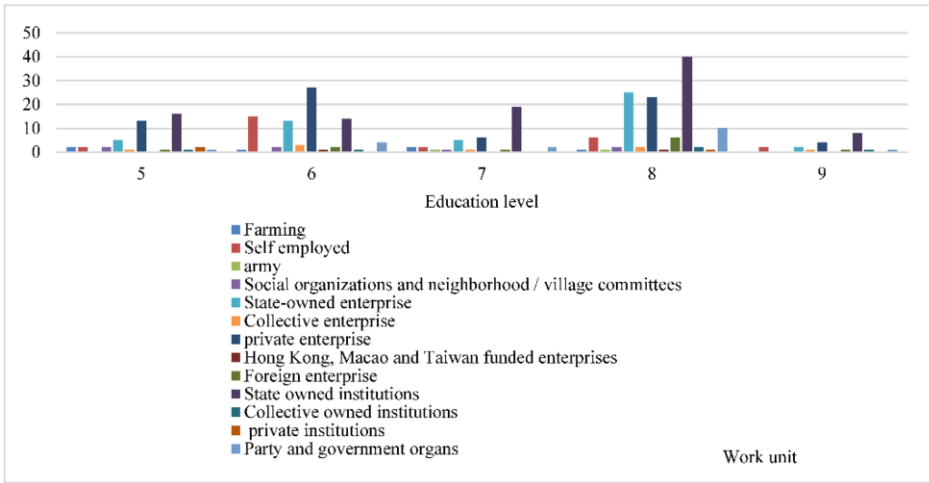


Figure 5. Cluster analysis of female samples in Higher Education

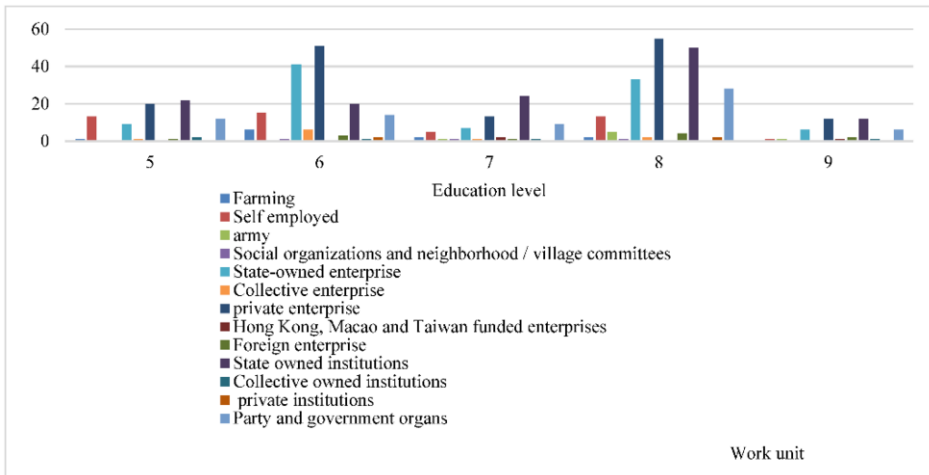


Figure 6. Cluster analysis of male samples in Higher Education

Among the female sample of higher education, more people are destined for state-owned institutions, state-owned enterprises, and private institutions. However, graduates of junior colleges (formal higher education) have outstanding performance in self-employment (including self-employed individuals) after graduation, which is far from other graduates. To a certain extent, this may indicate that women are more willing to go to stable state-owned units, which may be related to overall social expectations and long-term environment and education. Moreover, the graduates of colleges and universities are more adventurous and entrepreneurial, and the distribution of unit types in the other four education levels is basically similar.

In the sample of men with higher education, their employment tends to be more private enterprises, state-owned institutions and state-owned enterprises, but the distribution is more different than the sample of women. For example, the proportion of men of all educational levels who are going to be party and government agencies is

very high, and among college graduates and college graduates, there are many people who start their own businesses or do business. And it can be found that the total number of non-agricultural employment for males in college (formal higher education) and undergraduate (formal higher education) is similar, which is very different from the analysis of female samples. This may explain the tendency: in the long-term labor market, women's education requirements may be higher than men's. In contrast, women's overall employment orientation is relatively single. Under the influence of social culture and social expectation, for decades, women's choice of employment may be constantly restricted by the outside world and themselves, which together lead to the cycle of gender discrimination in work units or self-discrimination within women groups.

4 Conclusion

We divide the study of women's social status and its influence into two parts.

First of all, we use ordered logit model to analyze the rating of women's self-social status and various factors, discuss the factors that may affect women's social status, summarize the nature of the impact and explain it from the perspective of society, family and individual. Through this analysis, we can further understand the social facts shown in the current data. we find that women are more and more isolated from the influence of family, and are more involved in their social roles. At the same time, they pay more attention to their own value and experience. They think that their social status is related to their expectation and optimism, not only the material conditions or whether they are rich or not. But I also find that for high-level women, there is still a part of self-internal sexism.

Secondly, as an exploratory research, based on the background of big data, we conduct data mining on the relationship between education level and employment destination between different genders, which reflects the social status of women in employment. Using the parallel K-means clustering algorithm based on MapReduce. In this process, we improve and optimize the classification algorithm, and finally has better clustering results for the existing gender and work unit data. We find out the different employment situation between men and women, and explain the distribution and phenomenon. we hold that the different employment trends between men and women are based on the long-standing social expectation and cultural influence. The social atmosphere is more expecting women to manage housework, and more expecting men to work. This different pressure on gender makes women more inclined to choose state-owned units no matter what level of education they are in, while men are more likely to germinate the idea of entrepreneurship because of the pressure of earning money. Therefore, the society should gradually form an atmosphere for individuals to choose their own employment, rather than a moral criticism of a gender group. This is also conducive to women and men to explore their self-awareness, while promoting social division of labor equality, family division of labor freedom. At the same time, we also find that gender discrimination in the labor market may exist for a long time, that is, the educational requirements for women may be higher. To some extent, this affects the judgment of women's social status. But with the continuous implementation of the policy, such a phenomenon will inevitably be weakened.

In view of the current situation, the government should pay close attention to the gender dynamics, timely feedback to avoid the disorder of social order, and at the same

time, it should control the improper or cross-border speech in laws and regulations. Social media should be aware of the importance of self-guidance of public opinion and make objective reports and analysis. At present, with the development of network information, the public opinion guidance of various information channels may be tendentious, so that some people who do not have firm self-consciousness and self-worth are guided or even misled by them. Therefore, as a media, we should hold an objective and rational attitude all the time, and should not have strong emotional guidance. Each social platform should improve its own relevant laws and regulations, and abide by the implementation. The social platform should shoulder the huge social responsibility, standardize the use of rules, delete and manage the strong offensive or biased speech.

References:

- [1] Dixon R B. Rural women at work: strategies for development in South Asia. [J]. Baltimore, Maryland, Johns Hopkins University Press for Resources for the Future, 1978, 1978.
- [2] Chen Zaihua. Discussion on Comprehensive Evaluation Indicators of Women's Status [J]. China Population Science, 1993, 39 (6): 38-44.
- [3] Wei Huilan, Yang Yan. Study on evaluation index system of women's status [J]. Journal of Lanzhou University, 1999 (02): 97-103.
- [4] Ye Wenzhen, Liu Jianhua, Du Juan, Xia Yiran. Social status of Chinese women and its influencing factors [J]. Journal of population, 2003 (05): 22-28.
- [5] Liu Na. Application of MapReduce based data mining algorithm in national population system [D]. Capital University of economics and trade, 2011.
- [6] Huseyin Guruler, Ayhan Istanbulu, Mehmet Karahasan. A new student performance analysing system using knowledge discovery in higher educational databases[J]. Computers & Education, 2010, 55(1): 247-254.
- [7] Tao Tan, Leting Tan. Study on Personalization Recommendation System Based on Recruitment Information[J]. Procedia Engineering, 2012, 29: 780-784.
- [8] Ade R, Deshmukh P R. An incremental ensemble of classifiers as a technique for prediction of student's career choice[C]. First International Conference on Networks&Soft Computing (ICNSC), IEEE, 2014.
- [9] Parmar K, Vaghela D, Sharma P. Performance prediction of students using distributed data mining[C]. International Conference on Innovations in Information Embedded and Communication Systems (ICHECS), 2015: 1-5.
- [10] Qiao Fei, Ge Yanhao, Kong Weichang. Study on the classification model of distributed and improved random forest student employment data based on MapReduce [J]. System engineering theory and practice, 2017, 37 (05): 1383-1392.
- [11] Sun Xiaoyu, Zheng Yifang, Xu Jiaxian. An Empirical Study on the influencing factors of women's social status based on the comprehensive social survey data of China in 2013 and 2015 [J]. Journal of Shijiazhuang Railway University, 2018, 12 (04): 63-72.
- [12] William J. Scarborough, Danny L. Lambouths, Allyson L. Holbrook. Support of workplace diversity policies: The role of race, gender, and beliefs about inequality[J]. Social Science Research, 2019, 79.