

Classification of Land Production Function Based on Cluster Model

Xiao He, Kui Fang* and Xinghui Zhu

College of Information Science and Technology, Hunan Agricultural University, Changsha, China

*Corresponding author

Abstract—Scientific classification of land production functions can promote the efficient use of land. In this paper, 16 land production functions were collected, and 10 characteristic features of land production function were put forward according to the different attributes of each function. The characteristics of the land production functions are assigned to obtain the land production function score table. Then use the hierarchical clustering and K-means clustering analysis those data, and the results of three kinds of hierarchical clustering and one K-means clustering were obtained. By comparing with the results of sequence classification, 16 land production functions are classified into three categories.

Keywords—functions classification; cluster analysis; land

I. INTRODUCTION

Since the twenty-first century, Work on land use surveys at the regional and national levels is carried out in countries around the world. Opened the classification and evaluation of land functions and other basic research of land science. With the deepening of land resource classification research at China and other countries, there are many problems [1]. For example, most of the current research is based on the functional evaluation or value estimation of single target, or the experience of land use. The classification of land functions, has not yet formed a system of reference to the land functional classification system. At the more basic level, the research on the classification of land function is obviously lacking, and the system of land functional classification is not formed. Only through the land function classification research, to build an objective framework of the functional classification of land, it is possible to build a scientific and systematic land functional classification system.

Many domestic and foreign scholars have explored the preliminary classification of land function [2-3], Blum WE [4] base on the nature of soil function differences, divided the soil function into six parts. Banko [5] divided the land function into three main modules: the land economy function, the land social function and the land ecological function. Groot [6] summarized the land function into four broad categories, adding land regulation function and land information function beside on the basis of land production and land social function. Baotong Zhou [7] believe that the land function can be composed of five parts, land ecological function, land social function, land production function, land economic function and land negative function. Jing Chen [8] proposed that the land has three main functions which are production, ecology and life, and pointed out that the functions of the land are antagonistic and unified relations, especially highlighting the intrinsic relationship between the

land functions. Pei Liu [9] proposed that the land has three functions of production, society and ecology. Each function takes the main function of the land as the direct division basis, and establishes the secondary classification of the land function. This classification method is in reality has a good operability, the land function classification system has a guiding significance. However, the above methods are artificial classification, lack of scientific analysis, this paper presents the characteristics of land production function factors, and through the Hierarchical clustering and K-means clustering analysis, to get the land production function classification results.

II. LAND PRODUCTION FUNCTION TYPE AND CHARACTERISTIC FACTOR SETTING

According to the results of the scholars' research and the situation of land use classification promulgated by some countries, the social function of the land is divided into food function, economic crop function, other crop function, artificial forest function, natural forest function, forage function, artificial grass function, other grass function, livestock function, wildlife function, other biological production function, precious metal function, common metal function, rare earth function, energy function and building materials function, a total of 16 functions.

On this basis, there are two types of production location, surface production and underground production, surface production refers to the main production sites on the ground and shallow soil, underground production refers to the main production sites in the deep soil and below. There are two modes of production methods, manual intervention and natural production, manual intervention is meant to be carried out under human production activities. There are four types of output, animals, plants, metal minerals and nonmetallic minerals, in which animals are the output of animal type, plants refers to the output of plant type, metal mineral is the type of mineral production, nonmetallic minerals are the outputs of nonmetallic minerals. According to the nature of the land production function itself, as well as the definition of the characteristic factors, take yes or no to evaluate the value, the function has the characteristics of this feature factor assignment 1, the function doesn't have this feature factor that is assigned 0. Specific analysis of each land production function can be obtained as shown in Table I.

TABLE I. SCORE OF LAND PRODUCTION FUNCTION

Function name	Location of production		Methods of production		The type of output			
	Surface production	Underground production	Manual intervention	Natural production	Animals	Plants	Metal minerals	Nonmetallic minerals
Food function	1	0	1	0	0	1	0	0
Economic crop function	1	0	1	0	0	1	0	0
Other crop function	1	0	1	0	0	1	0	0
Artificial forest function	1	0	1	0	0	1	0	0
Natural forest function	1	0	0	1	0	1	0	0
Forage function	1	0	1	0	0	1	0	0
Artificial grass function	1	0	1	0	0	1	0	0
Other grass function	1	0	0	1	0	1	0	0
Livestock function	1	0	1	0	1	0	0	0
Wildlife function	1	0	0	1	1	0	0	0
Other biological production function	1	1	0	1	1	0	0	0
Precious metal function	0	1	0	1	0	0	1	0
Common metal function	0	1	0	1	0	0	1	0
Rare earth function	0	1	0	1	0	0	0	1
Energy function	0	1	0	1	0	0	0	1
Building materials function	0	1	0	1	0	0	0	1

III. CLUSTERING RESULTS AND ANALYSIS OF LAND PRODUCTION FUNCTION

A. Hierarchical Clustering and K-means Clustering Results

The data of Table 1 were clustered by the Ward minimum variance method hierarchical clustering and the K-means clustering with the maximum number of iterations of 10. As the number of land production functions are 16, regardless of the method used to be divided into 16 categories are each one to form a separate class, no matter what method to use it into a class when all 16 items into a large category, so the results in the direct omitted. Use SPSS (23.0) statistical analysis of software operations, to obtain the results for follow-up analysis.

B. Analysis of the Optimal Clustering Number of Genealogy

In the clustering analysis itself can not automatically determine the optimal number of clusters, so the need for the classification of land production function of the results to find out the best clustering analysis, select the Semipartial R-Square statistics, R-Square statistics, Pseudo F statistics and Pseudo t-Squared statistics, those 4 statistics often used to find the best clustering categories, through SAS (9.4) statistical analysis

software operation, by use of excel to production results shown in Figure I, Figure II, Figure III, Figure IV.

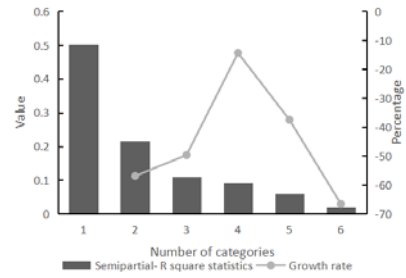


FIGURE I. SEMIPARTIAL R-SQUARE STATISTICS

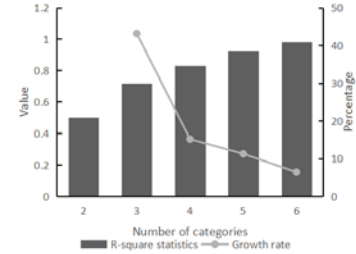


FIGURE II. R-SQUARE STATISTIC

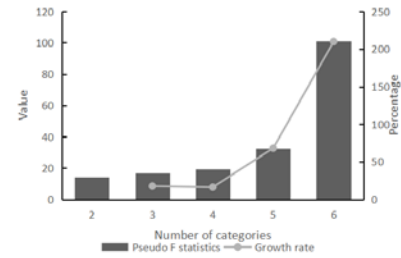


FIGURE III. PSEUDO F STATISTICS

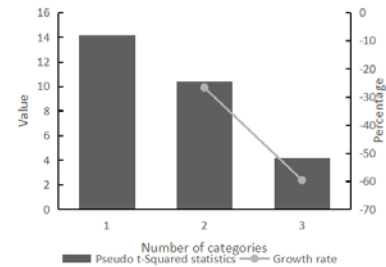


FIGURE IV. PSEUDO T-SQUARED STATISTICS

It can be seen from Figure I that the value of the semi-partial R2 statistic decreases gradually with the number increase, and its growth rate has four relatively low values, respectively, while the number of categories is 2, 3, 5, 6, The growth rate is between -30% to -70%, according to the definition of the statistics, should be selected the larger number of 5 or 6 of the four items as the best classification number, but the value of the semi-partial R2 statistic when the number of classifiers is 5 or 6 is less than the value of 4 or 5, which is reduced from 0.0934 to 0.0584 and 0.0584 to 0.0195, respectively, Is 0.04, only because the base number is too low, it makes the growth rate is low, so

according to the definition of semi-partial R2 statistics, do not select 5 or 6 as the best classification number. When the number of categories is 2 or 3 the growth rate is lower than the number of categories is 1 or 2, so choose 2 and 3 as the best classification of the statistics. It can be seen from Figure II that the value of R2 statistic increases gradually with the number increases, and its growth rate is significantly slowed down at the beginning of the number of classes 4 and maintained in a lower interval, So according to the definition of the statistics selected 3 as the statistics of the best classification number. It can be seen from Figure III that the value of the pseudo-F statistic increases gradually with the increasing number increases, and the growth rate increases gradually. Therefore, the optimal classification number can not be selected according to the definition of the statistics. It can be seen from Figure IV that the value of the pseudo t2 statistic does not have a local maximum value. Therefore, the best classification number can not be selected according to the definition of the statistic.

C. K - means Clustering Optimal Clustering Analysis

Since the clustering process of K-means clustering is not a simple merging grouping,we can not calculate the semi-partial R2 and pseudo t2 statistic.We choose R2 and pseudo-F statistic to optimize the clustering results of K-means clustering.Use SAS(9.4)statistical analysis software operation,use excel to produce the results shown in Figure V,Figure VI.

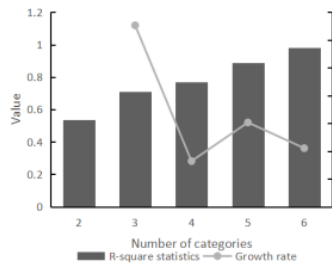


FIGURE V. R-SQUARE STATISTICS

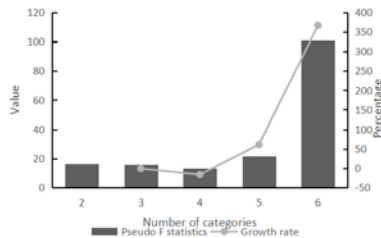


FIGURE VI. PSEUDO F STATISTICS

It can be seen from Figure V that the value of R2 statistic increases gradually with the number increases, and its growth rate is higher when the number of classification items is 3, and the value of R2 statistic is larger when the number of classification is 3, Then according to the definition of the statistics selected 3 as the statistics of the best classification number. It can be seen from Figure VI that the value of the pseudo-F statistic fluctuates slightly in the interval during the number of classifications 2-5, and the value of the pseudo-F statistic corresponding to the number of classifiers increased rapidly, and does not have a local maximum, so according to the definition of the statistics can not choose the best number of categories.

D. Classification Result Analysis and Analysis

Based on the results of the optimal classification number of the hierarchical cluster and K-means clustering,it can be concluded that the 2 categories and 3 categories of the hierarchical cluster,and the 3 categories of the K-means clustering are the best classification number recommended by those statistics.And then compared with the sequential classification obtained by querying the literature.The results are shown in Table II.

TABLE II. COMPARISON OF CLUSTERING RESULTS

Function name	Hierarchical clustering		K-means clustering	Sequential classification
	2 Classes	3Classes	3 Classes	2 Classes
Food function	1	1	1	1
Economic crop function	1	1	1	1
Other crop function	1	1	1	1
Artificial forest function	1	1	1	1
Natural forest function	1	2	1	1
Forage function	1	1	1	1
Artificial grass function	1	1	1	1
Other grass function	1	2	1	1
Livestock function	1	2	1	1
Wildlife function	1	2	2	1
Other biological production function	1	2	2	1
Precious metal function	2	3	3	2
Common metal function	2	3	3	2
Rare earth function	2	3	3	2
Energy function	2	3	3	2
Building materials function	2	3	3	2

First of all, it can be found that the hierarchical clustering divide the land production function into two categories are in good agreement with the sequential classification,which is divided into two categories:biological production function and mineral production function. Food function, economic crop function, other crop functions, artificial forest function, natural forest function, forage function, artificial grass function, other grass function,livestock function,wildlife function,other biological production function, and the other contains precious metal function,ordinary metal function,rare earth function, energy function, building materials function, this classification target is relatively clear, according to the nature of the difference between the characteristics of classification, the results are more concise, but the classification is not enough refinement. Secondly, we can find that K-means clustering divides the land production function into three categories, which is based on the results of hierarchical clustering and sequential classification .The wild animal function and the other biological production function is divided into a single class, and this single class can not form a parallel relationship with the other two categories, so the comprehensive view, this classification is not sufficient reason, only a single refinement of a class classification, the results not ideal. Then, it is found that when

the hierarchical cluster is divided into three categories is divided the natural forest function, Function, livestock function, wildlife function, other animal production function and other functions into a single class. This classification seems to be divided natural productive functions alone into a class, but in turn the apparent function of human intervention in the livestock function, although the results are not ideal, but this classification provides more detailed classification thinking. Based on the above analysis, we can sort out the classification results in Table III.

TABLE III. CLASSIFICATION RESULT

Category name	Function name
Artificial biological production function	Food function
	Economic crop function
	Other crop function
	Artificial forest function
	Livestock function
	Forage function
Natural biological production function	Artificial grass function
	Natural forest function
	Other grass function
	Wildlife function
Mineral production function	Other biological production function
	Precious metal function
	Common metal function
	Rare earth function
	Energy function
	Building materials function

All the functions are divided into three categories: artificial biological production function, natural biological production function, mineral production function, combined with the advantages of hierarchical clustering concise and clear, but also continued the k-means clustering and hierarchical clustering refinement classification of ideas, in general, clear classification objectives, the results clear.

IV. SUMMARY

This paper compared the method of hierarchical clustering, k-means clustering and sequential classification, which laid a foundation for the further study of the classification of land production function. However, this study has a few shortcomings: only two kinds of clustering methods, a distance algorithm and four kinds of validity analysis indexes are selected because of the limitation of research conditions. Therefore, we are only an initial exploratory study of the classification of land production functions.

The classification of land function is a long-term basic work, the classification of land function is worth further study. In the future, we can use more types of clustering analysis and neural network technology, more in-depth classification of research, to promote the further improvement of land functional classification system.

ACKNOWLEDGMENT

This paper is supported by the research and development key projects in Hunan Province (No 2016NK2118) and national R & D key projects (No 2017YFD0301506).

REFERENCES

- [1] Hietel E, Waldhardt R, Otte A. Analysing land-cover changes in relation to environmental variables in Hesse, Germany[J]. *Landscape Ecology*, 2004, 19(5):473-489.
- [2] B M Chen, X P Zhou. Explanation of current land uses condition classification for national standard of people's republic of China[J]. *Journal of Natural Resources*, 2007, 22 (6): 994-1003.
- [3] Latham J S, He C, Alinovi L, et al. FAO Methodologies for Land Cover Classification and Mapping[M]// *Linking People, Place, and Policy*. 2002:283-316.
- [4] Blum W E , Lal R, Blum W H, et al. Basic concepts: degradation, resilience, and rehabilitation.[M]. 1998.
- [5] Banko G, Mansberger R. ASSESSMENT OF 'NON MONETARY VALUES OF LAND' FOR NATURAL RESOURCE MANAGEMENT USING SPATIAL INDICATORS[C]// 2001:2--5.
- [6] Groot R D, Groot R D. Function-analysis and valuation as a tool to assess land use conflicts in planning for sustainable, multi-functional landscapes[J]. *Landscape & Urban Planning*, 2006, 75(3-4):175-186.
- [7] B T Zhou. Study on basic theoretical concepts of sustainable land uses[J]. *Journal of Southwest China Normal University (Natural Science Edition)*, 2004, 29 (2): 310-314.
- [8] Q Chen, P J Shi .Discussion on functional land uses classification system [J]. *Journal of Beijing Normal University (Natural Science)*, 2005 (5): 536-540.
- [9] P Liu, J N Duan, W Wang, et al. Study on Systems of the land-uses system functional classification and evaluation [J]. *Journal of Hunan Agricultural University (Special Edition)*, 2010, 36 (1): 113-118.