Research on the Application of Network Planning Technique in Big Data Environment

Zhi-hang WANG* and Xue-mei LI
Beijing Jiaotong University, Beijing, China
*Corresponding author

Keywords: Big data, Network planning technique, Three-point time estimation, Cluster analysis.

Abstract. As an important part of scientific management, network planning technique has a very wide range of applications. In big data environment, network planning technique still has a significant role. The estimation of the activity time is the important aspects of the network planning technique. The three-point time estimation method is a commonly used method to estimate the time of activity. However, there are many shortcomings in the estimation method. On the basis of the traditional estimation method and the big data environment, this paper proposes a method based on cluster analysis. In this paper, the method of estimating the time of activity in the big data environment are beneficial to estimate the activity time so as to shorten the construction period.

Introduction

With the development of the economy, the living standards of people have been continuously improved, which has led to the increasing demand of consumers for specialization. The change in consumer values has also led to certain changes in the production mode of enterprises. In order to ensure their own competitive advantages and gain more profits, corporations have gradually transformed from the original small batch production model to the multi-variety small batch production model[1].

As an important method of project planning and management, network planning technique has become more and more widely used. In big data environment, more emphasis should be placed on its application. Working time is a major focus of corporate competition. In an increasingly competitive environment, companies are working to shorten the construction period and reduce costs. The determination of the activity time is an important part of the network planning technique. In the big data environment, it is more convenient to obtain data, providing a large amount of data support for this research. The determination of the time for the activity will be more objective and reasonable, and it will provide certain guidance for the production and implementation of the company, thereby shortening the construction period and reducing the cost. The three-point time estimation method is a commonly used method for estimating activity time. Scholars gradually realized the shortcomings of the traditional three-time estimation method and devoted to studying new estimation methods based on traditional methods to make up for the deficiencies of traditional methods. Zhang Yunning and Sun Xixi (2008) elaborated the shortcomings of the traditional three-time estimation method, proposed a corresponding method to improve the estimation method of the PERT project duration[2]. Cheng Chao, Du Zhiyuan et al. (2008) expounded the shortcomings of the traditional three-time estimation method, and proposed using the limited probability three-time estimation method to establish the Beta and the distribution variance of Gamma distribution respectively. The minimum model uses the particle swarm algorithm to solve the model and compares the results with the traditional three-time estimation method [3]. In big data environment, this paper proposes the determination of activity time based on clustering analysis. Using K-means cluster analysis and hierarchical clustering analysis, the original data is clustered into three categories, and the final cluster center replaces the most optimistic time, the most probable time and the most pessimistic time in the big data environment. Finally determine the time of the activity on the basis of traditional methods.
Related Theory

Three-Point Time Estimation

In the traditional planning review technology, the three-point time estimation method is the most commonly used method to estimate the activity time. This method is mainly applicable to exploratory engineering projects. The network diagram drawn by this method is a random network diagram. The three-point time estimation method first estimates three time values for various activities, and then obtains the weighted average value according to the formula to obtain the activity time $t(i,j)$.

$$t(i, j) = \frac{a + 4m + b}{6}$$  \hspace{1cm} (1)

Among them, $a$ represents the most optimistic time, the time needed to complete an activity under the most favorable conditions; $m$ represents the most probable time, the time required to complete an activity under normal conditions; $b$ represents the most pessimistic time, the time required to complete an activity under the worst conditions.

Cluster Analysis

Cluster analysis is to classify data according to certain conditions, divide the original and more data into several related data, which is conducive to understand and study the relationship between these data. The method is to classify the collected data into different clusters. After cluster analysis, the data in the same cluster has a great similarity, and the data of different clusters have great differences.

The main steps of cluster analysis are divided into the following steps:

1. Feature extraction: that is, which feature is selected for the classification and extraction of the original data.
2. Obtain the clustering pedigree map: After the original data is subjected to feature extraction, a matrix with features can be formed, and then the clustering algorithm is used to obtain the cluster pedigree map.
3. Determine the appropriate classification threshold: According to the cluster pedigree obtained in the previous step and the specific purpose of the classification, select the appropriate threshold.

Time Estimation Based on Cluster Analysis

The commonly used methods of cluster analysis are mainly classified into the following categories: partition method, hierarchical method, density-based method, and grid-based method. We can choose different cluster analysis methods according to different classification purposes, and the K-means algorithm is a commonly used clustering algorithm for the division method. This algorithm uses distance similarity as the evaluation criterion, and its purpose is to divide the closest raw data into the same cluster. As a common software tool for cluster analysis, SPSS can perform three kinds of cluster analysis, such as two-step cluster analysis, rapid cluster analysis and hierarchical cluster analysis, and can play a role in data mining. This paper mainly uses the K-means algorithm and hierarchical clustering analysis to solve the problem of determining the activity time in the big data environment.

K-means Clustering

Using K-means algorithm for clustering, the activity time is divided into three categories: the most pessimistic time, the most probable time and the most optimistic time.

1. Preparation phase: Prepare the raw data of the activity time collected before, open it in the SPSS software, and open the K-means clustering dialog according to “Analysis - Classification - K-means Clustering”;
2. The number of clusters is determined in advance according to specific needs, so K-means clustering results in a unique solution. This is based on the traditional three-point time estimation...
method and proposes a method for determining the activity time in the big data environment. Therefore, the collected raw data should be grouped into three categories:

(3) Set the requirements according to the steps, run the data and obtain the result;

(4) Analyze the results and obtain the corresponding information.

The process of determining the activity time based on the K-means clustering analysis will be described below through a set of data. Suppose a set of data collected for a certain activity time is as follows: 4.5, 9.1, 2, 3, 3, 4, 5, 6, 7.5, 6.5, 2.8, 3.5, 4.8, 5.8, 8.5, 15.1, 15.5, 15, 2.5, 8.3, 8.2, 4.8, 10, 10.7, 12, 12.5, 10.1, 13, 13.2, 10.9, 13.5, 13.7, 6.6, 8.8, 3.8, 9, 7, 19, 15, 14.8, 14, 18.5, 15.3, 15, 15.9, 18, 13.9, 14.5, 11.5, 14.7, 16, 9.5, 9.8, 11.2, 9.9, 17, 18.8, 9.6, 10, 10.5, 11.8, 11.7, 10.2, 11, 12, 10.8, 12.3, 5.6, 6.8(unit: min).

First, this group of data is entered into the SPSS software, and the K-means dialog box is opened. The requirements are set in order and the following results are finally obtained. Table 1 shows the initial clustering center. After thirteen iterative processes in Table 2, the final clustering result is obtained.

| Activity time | 1   | 2   | 3   |
|---------------|-----|-----|-----|
|               | 10.5| 19  | 2   |

Table 2. The Iteration History

| Iteration | Cluster Center Changes |  |  |
|-----------|------------------------|---|---|
| 1         | 0.519                  | 3.417 | 2.25 |
| 2         | 0.272                  | 0.403 | 0.006 |
| 3         | 0                      | 0.016 | 0   |
| 4         | 0                      | 0.001 | 2.04E-05 |
| 5         | 9.14E-06               | 2.58E-05 | 1.20E-06 |
| 6         | 2.95E-07               | 1.03E-06 | 7.04E-08 |
| 7         | 9.51E-09               | 4.13E-08 | 4.14E-09 |
| 8         | 3.07E-10               | 1.65E-09 | 2.44E-10 |
| 9         | 9.89E-12               | 6.61E-11 | 1.43E-11 |
| 10        | 3.27E-13               | 2.64E-12 | 8.42E-13 |
| 11        | 8.88E-15               | 1.05E-13 | 5.06E-14 |
| 12        | 0                      | 5.33E-15 | 2.67E-15 |
| 13        | 0                      | 0      | 0   |

After a series of operations, the original data was finally gathered into three categories as required. The first category is the most likely time group: 7, 7.5, 8, 8.2, 8.3, 8.5, 8.8, 9, 9.1, 9.5, 9.6, 9.8, 9.9, 10, 10.1, 10.2, 10.5, 10.7, 10.8, 10.9, 11, 11.2, 11.5, 11.7, 11.8, 12, 12, 12.3,a total of 29 data; the second is the most pessimistic time group: 12.5, 13, 13.2, 13.5, 13.7, 13.9, 14, 14.5, 14.7, 14.8, 15, 15, 15.1, 15.3, 15.5, 15.9, 16, 17, 18, 18.5, 18.8, 19, a total of 23 data; the third is the most optimistic time group: 2, 2.5, 2.8, 3, 3, 3.5, 3.8, 4, 4.5, 4.8, 4.8, 5, 5.6, 5.8, 6, 6.5, 6.6, 6.8, contains a total of 18 data. From Table 3, the final cluster centers for the most optimistic time groups, most probable time groups, and most pessimistic time groups are 4.5, 10, and 15.3, respectively.

| Activity time | 1   | 2   | 3   |
|---------------|-----|-----|-----|
|               | 10  | 15.3| 4.5 |
Similarly, using this method to process the same data collected, using the final cluster center as the most optimistic time, the most probable time and the most pessimistic time for activities under the big data environment, and then according to the traditional three points The time estimation method determines the operation time of each activity.

Hierarchical Clustering Analysis

Hierarchical clustering is the hierarchical decomposition of collected data according to a certain method, until it reaches a certain condition. The basic principle is: First, each object is divided into a class, n objects are divided into n classes, then the distance between the class and the class is the distance they contain the object, and then merge the two closest classes into One class, and recalculate the distance between the class and the class, and repeat the second step until the result is obtained.

The above data is still taken as an example to introduce the application of hierarchical cluster analysis in determining the active working time in a big data environment. First of all, this group of data is entered into the SPSS software, and the hierarchical clustering dialog box is opened. The requirements are set in order and the results of hierarchical clustering can be finally obtained.

In this example, the method of clustering within the group is adopted. After a series of iterative processes, the collected raw data can be finally gathered into three categories. Table 4 shows the result of clustering the data into 3 categories. From the table, It can be seen directly which specific data is included in each category. Reading the table shows that the first cluster is the most optimistic time: 2, 2.5, 2.8, 3, 3.5, 3.8, 4, 4.5, 4.8, 4.8, 5, 5.6, 5.8, 6, 6.5, 6.6, 6.8, 7, A total of 19 items of data;  the second cluster is the most likely time: 7.5, 8, 8.2, 8.3, 8.5, 8.8, 9, 9.1, 9.5, 9.6, 9.8, 9.9, 10, 10, 10, 10, 10, 10, 10, 10, 10.5, 10.7, 10.8, 10.9, 11, 11.2, 11.5, 11.7, 11.8, 12, 12, 12, 12, 12.5, a total of 29 data; the third cluster is the most pessimistic time: 13, 13.2, 13.5, 13.7, 13.9, 14, 14.5, 14.7, 14.8, 15, 15, 15, 15.1, 15.3, 15.5, 15.9, 16, 17, 18, 18.5, 18.8, 19. A total of 22 data are contained.

| Case | Cluster | Case | Cluster | Case | Cluster | Case | Cluster |
|------|---------|------|---------|------|---------|------|---------|
| Case 1 | 1       | Case 15 | 2       | Case 29 | 3       | Case 43 | 3       | Case 57 | 3       |
| Case 2 | 2       | Case 16 | 3       | Case 30 | 3       | Case 44 | 3       | Case 58 | 3       |
| Case 3 | 1       | Case 17 | 3       | Case 31 | 2       | Case 45 | 3       | Case 59 | 2       |
| Case 4 | 1       | Case 18 | 3       | Case 32 | 3       | Case 46 | 3       | Case 60 | 2       |
| Case 5 | 1       | Case 19 | 1       | Case 33 | 3       | Case 47 | 3       | Case 61 | 2       |
| Case 6 | 1       | Case 20 | 2       | Case 34 | 1       | Case 48 | 3       | Case 62 | 2       |
| Case 7 | 1       | Case 21 | 2       | Case 35 | 2       | Case 49 | 3       | Case 63 | 2       |
| Case 8 | 1       | Case 22 | 2       | Case 36 | 1       | Case 50 | 2       | Case 64 | 2       |
| Case 9 | 2       | Case 23 | 1       | Case 37 | 2       | Case 51 | 3       | Case 65 | 2       |
| Case 10 | 1      | Case 24 | 2       | Case 38 | 1       | Case 52 | 3       | Case 66 | 2       |
| Case 11 | 1      | Case 25 | 2       | Case 39 | 3       | Case 53 | 2       | Case 67 | 2       |
| Case 12 | 1      | Case 26 | 2       | Case 40 | 3       | Case 54 | 2       | Case 68 | 2       |
| Case 13 | 1      | Case 27 | 2       | Case 41 | 3       | Case 55 | 2       | Case 69 | 1       |
| Case 14 | 1      | Case 28 | 2       | Case 42 | 3       | Case 56 | 2       | Case 70 | 1       |

It can be seen from the above clustering results that the results obtained by the two different clustering methods are basically the same, which indicates that the clustering analysis has certain application value in determining the working time of the activity. In the big data environment, this method can be used to process the collected large amount of data, using the final cluster center as the most optimistic time, the most likely time and the most pessimistic time of the activity in the big data environment, and then determines the working time of each activity according to the traditional three-point time estimation method.
Summary

Based on the knowledge of network planning technology and the related theories found in other aspects, this paper introduces the determination of operating time based on activities in the context of big data based on traditional methods.

The main results of this paper are the following:

1. Explained the traditional method of estimating activity time—the insufficiency of the three-point time estimation method, and pointed out that the estimated activity time of the method is too subjective;

2. The method of determining the working time based on the activity in big data environment is proposed. This paper mainly introduces the method of determining the active working time based on clustering analysis.

Through the research of this paper, the estimation of activity time is more accurate than the traditional method, and it has certain guiding significance for the production of the enterprise. Due to the limitations of knowledge and practical experience, the paper only gives several simple methods for determining the activity time in the big data environment. Although these methods have made great progress compared with the traditional methods, but these methods also have some problems. For example, if the sample is large, there are certain difficulties in obtaining clustering conclusions. Future scholars will continue to do more in-depth research in this area, looking for more reasonable ways to make up for the deficiencies.

Acknowledgement

First and foremost, I would like to show my deepest gratitude to my supervisor, Dr. Li Xuemei, a respectable, responsible and resourceful scholar, who has provided me with valuable guidance in every stage of the writing of this paper. Without her enlightening instruction, impressive kindness and patience, I could not have completed my paper. I shall extend my thanks to Mrs. Li for all her kindness and help. I would also like to thank all my teachers who have helped me to develop the fundamental and essential academic competence. Last but not least, I'd like to thank all my friends, for their encouragement and support.

References

[1] J. Yang Hongying. On the New Mode of Multi-variety and Small-batch Production Management, SME Management and Technology.09(2013)59-60.

[2] J. Zhang Yunning, Sun Xixi. Application of Improved PERT Model in Project Schedule Risk Analysis, Technological Progress and Countermeasures.10(2008)94-96.

[3] J. Cheng Chao, Du Zhida, Chen Xiaoyan. Study on the Improvement of Activity Time Parameters in Uncertain Networks by Using PERT Method, Value Engineering.12(2008)104-108.

[4] W. Chen Rongqiu, Ma Shihua. Production Operations Management. Beijing: Machinery Industry Press, 2015.

[5] J. Sun Yonghui. Application of Cluster Analysis in Student Performance Analysis, China Management Information.06(2016)229-230.