Research on the Improvement of Airport Comprehensive Support Capability Based on Data Mining

Xiuming Zhang*, Jianhai Xue*
Guangzhou Civil Aviation College, Guangzhou, 510410, China

*Corresponding author e-mail: xuejianhai@caac.net, zhangxiuming@caac.net

Abstract. The "integrated airport support capability" is a general term for all the ground support capabilities in the airport relative to the sky. For the protection of the airport, especially during peak hours, it is necessary for multiple units to coordinate and participate together. The best guarantee of the apron in an airport is the operational efficiency during peak hours. This paper firstly identifies the support capacity of an airport during peak hours, and then analyzes the data of the airport peak period, so it is concluded that those factors play an important role in the key guarantee of the airport peak and the key guarantee for the airport. The factors put forward the opinions of optimization, which can greatly enhance the comprehensive support capability of the airport.

1. Introduction

“A comprehensive airport support capability” is a relatively broad concept, and it is a general term for all ground support capabilities relative to the opposite side of the sky. Broadly speaking, the ground support capability refers to the ability of airports, airlines and their agencies to provide various guarantees for passengers and cargo owners, as well as air traffic control, aircraft maintenance companies, and aviation fuel companies to provide services to airlines. Ability. The comprehensive airport support capability studied in this paper refers to various ground support operations provided by aircraft, aircraft, airlines, or agencies and other units to aircraft during arrival and departure.

The main work of the airport comprehensive guarantee is to arrange the airport ground support vehicles reasonably under the premise of ensuring safety, and to ensure the maintenance of the aircraft in an orderly manner in the shortest possible time, to avoid flight delays as much as possible, and the ground support work is often subject to flight schedules. The impact of the passenger capacity and cargo capacity of the flight, the project for the flight receiving service, the aircraft model, and the service route of the ground support vehicle. For airport security, especially during peak hours, it is necessary for all parties to coordinate and participate. In the development of modern airports, it is necessary to improve the operational capacity under the premise of safety, and the most effective guarantee capability of the apron is the operational efficiency during peak hours. This paper analyzes the peak hours of an airport to obtain key focus factors, and aims to improve the level of airport security.

2. Analysis of airport apron support capabilities

“A comprehensive airport support capability” is a relatively broad concept, and it is a general term for all ground support capabilities relative to the side of the sky. Broadly speaking, the ground support
capability refers to the ability of airports, airlines and their agencies to provide various guarantees for passengers and cargo owners, as well as air traffic control, aircraft maintenance companies, and aviation fuel companies to provide services to airlines. Ability. The comprehensive airport support capability studied in this paper refers to various ground support operations provided by aircraft, aircraft, airlines, or agencies and other units to aircraft during arrival and departure.

The main work of the airport comprehensive guarantee is to arrange the airport ground support vehicles reasonably under the premise of ensuring safety, and to ensure the maintenance of the aircraft in an orderly manner in the shortest possible time, to avoid flight delays as much as possible, and the ground support work is often subject to flight schedules. The impact of the passenger capacity and cargo capacity of the flight, the project for the flight receiving service, the aircraft model, and the service route of the ground support vehicle. For airport security, especially during peak hours, it is necessary for all parties to coordinate and participate. In the development of modern airports, it is necessary to improve the operational capacity under the premise of safety, and the most effective guarantee capability of the apron is the operational efficiency during peak hours. This paper analyzes the peak hours of an airport to obtain key focus factors, and aims to improve the level of airport security.

3. The application of data mining in the apron data

3.1. Application of data mining method in airport ground support data

In terms of improving the support capacity of the airport, it is theoretically necessary to strengthen and enhance all aspects of security. However, considering the reality, it is often necessary to protect the key protection content or the weak content to save costs. In order to solve the problem in a targeted manner, it is necessary to clarify the problem to be solved. There are many ways to find a problem, and there is an old and modern way of data mining. Broadly speaking, data mining is the extraction of information from the database. In terms of computer technology, data mining refers to the process of automatically analyzing and extracting information from data using one or more computer learning techniques. Data mining can extract hidden, unknown, but potentially useful information and knowledge from a large number of incomplete, noisy, fuzzy, and random real data.

3.2. Principle of K-means algorithm

The analysis method for peak hour data uses clustering method. The goal is to find an aggregation of related items. Aggregation is the grouping of records, with similar records in a single aggregate. The difference between aggregation and classification is that aggregation does not depend on pre-defined classes, and it is possible to gather similar data together without a training set. This step is generally the first step in data analysis. Cluster analysis means the process of dividing several objects without obvious classification features into multiple clusters according to some similarity. In the development of data mining, there are many clustering algorithms and techniques.

Among these methods, the most classic, most widely used K-means algorithm with good clustering effect. The K-means algorithm is a simple and effective clustering statistical technique proposed by Stuart Lloyd in 1982. Its basic content and principles are as follows:

1. Randomly select a K value to determine the total number of clusters.
2. Select K elements arbitrarily in the original data set and use the selected data as the original cluster center.
3. Calculate the Euclidean Distance between the K cluster centers and other remaining instances, using this distance as a measure of similarity between instances, with similar clusters as members.
4. Use the instances in each cluster to calculate the new cluster center for the cluster.
5. If the calculation calculates that the new cluster center is equal to the cluster center of the last iteration, the algorithm is aborted. Otherwise, repeat steps 3-5 with the new cluster center.

To illustrate the implementation flow of the K-means algorithm, let's take a brief example:

The points in this two-dimensional space are the original data, as shown in the following figure:
After several iterations, the following clusters are obtained:

![Raw data graph](image1)

**Figure 1.** Raw data graph

![Cluster final rendering](image2)

**Figure 2.** Cluster final rendering

3.3. Excavation of key projects for ground operation support of an airport

3.3.1. Airport ground operation support data. After clarifying the content process of the airport guarantee, from the perspective of operational support, the data of the airport's support characteristics are collated and analyzed to find the associated weaknesses of the airport's operational support. It is proposed to select the operational data of a certain quarter of an airport for data mining, and the selection of data should be a peak period in which the guarantee capability is obvious.

In order to understand the relationship between passenger throughput and annual throughput in the peak season in recent years, according to the data, the flight data of a large airport from July to September of 2014 to 2016 is shown in Table 1.


Table 1. Relationship between passenger throughput and annual throughput of an airport during summer vacation

| Basic data (year) | Passenger throughput from July to September | Annual passenger throughput | Annual ratio |
|-------------------|-------------------------------------------|----------------------------|--------------|
| 2014              | 1494226                                   | 4331529                    | 34.5%        |
| 2015              | 2028357                                   | 5435237                    | 37.3%        |
| 2016              | 2189054                                   | 6150282                    | 35.6%        |

From the data sheet, it can be intuitively found that during the July-September summer holiday period, it is also a time period when the guarantee capability is tested. According to the data, the statistics of the flight schedules during the peak hours during the summer vacation are as follows:

“A comprehensive airport support capability” is a relatively broad concept, and it is a general term for all ground support capabilities relative to the side of the sky. Broadly speaking, the ground support capability refers to the ability of airports, airlines and their agencies to provide various guarantees for passengers and cargo owners, as well as air traffic control, aircraft maintenance companies, and aviation fuel companies to provide services to airlines. Ability. The comprehensive airport support capability studied in this paper refers to various ground support operations provided by aircraft, aircraft, airlines, or agencies and other units to aircraft during arrival and departure.

The main work of the airport comprehensive guarantee is to arrange the airport ground support vehicles reasonably under the premise of ensuring safety, and to ensure the maintenance of the aircraft in an orderly manner in the shortest possible time, to avoid flight delays as much as possible, and the ground support work is often subject to flight schedules. The impact of the passenger capacity and cargo capacity of the flight, the project for the flight receiving service, the aircraft model, and the service route of the ground support vehicle. For airport security, especially during peak hours, it is necessary for all parties to coordinate and participate. In the development of modern airports, it is necessary to improve the operational capacity under the premise of safety, and the most effective guarantee capability of the apron is the operational efficiency during peak hours. This paper analyzes the peak hours of an airport to obtain key focus factors, and aims to improve the level of airport security.

Figure 3. Histogram and line chart of the takeoff and landing of the peak period

After sorting out the flight type and synthesizing the positioning of an airport, it will sort out the inbound and outbound flights. Finally, the next step is to process the inbound and outbound flight data sheets. The collated data table is as follows:
### Table 2. Flight Entry and Exit Checklist

| Serial number | model | Inbound flight | Blocking time | Outbound flight | Voyage time | Departure time |
|---------------|-------|----------------|---------------|-----------------|-------------|----------------|
| 1             | A319  | CBJ5599        | 18:21         | CBJ5600         | 1hour       | 19:20          |
| 2             | A190  | CSN6928        | 18:30         | CSN6928         | 1hour       | 19:25          |
| 3             | A320  | CSZ9410        | 18:33         | CSZ9410         | 2hour       | 19:45          |
| 4             | A319  | CBJ5745        | 18:38         | CBJ5746         | 2.5hour     | 19:40          |
| 5             | A145  | GCR6468        | 18:43         | GCR6467         | 1.5hour     | 19:15          |
| 6             | A737  | CDG4717        | 18:47         | CDG4718         | 1hour       | 19:40          |
| 7             | A738  | CSH9137        | 18:52         | CSH9138         | 2.5hour     | 19:15          |
| 8             | A320  | CSN6663        | 18:56         | CSN6664         | 1hour       | 19:30          |
| 9             | E145  | GCR6552        | 19:00         | GCR7407         | 1hour       | 19:55          |
| 10            | A319  | CBJ5664        | 19:04         | CBJ5664         | 2.5hour     | 19:45          |
| 11            | E145  | GCR6405        | 19:08         | GCR6405         | 0.5hour     | 19:35          |
| 12            | B738  | CHH7043        | 19:12         | CHH7044         | 3hour       | 19:55          |
| 13            | A320  | CES5460        | 19:15         | CES5460         | 2.5hour     | 19:55          |
| 14            | B772  | CCA1685        | 19:19         | CCA1686         | 1hour       | 20:00          |
| 15            | A738  | CDG4981        | 19:23         | CDG4982         | 1.5hour     | 20:05          |
| 16            | A737  | CES7675        | 19:30         | CES7676         | 3hour       | 20:45          |
| 17            | A320  | CHH7175        | 19:36         | CHH7176         | 1hour       | 20:30          |
| 18            | A190  | GCR7439        | 19:40         | GCR7539         | 2hour       | 20:15          |
| 19            | A319  | CSC8724        | 19:45         | CSC8724         | 1hour       | 20:20          |
| 20            | A320  | CBJ5339        | 19:49         | CBJ5311         | 1.5hour     | 21:00          |
| 21            | A190  | MNG801         | 19:54         | MNG802          | 2hour       | 21:50          |
| 22            | A319  | CCA1792        | 19:57         | CCA1792         | 2hour       | 20:55          |

### 3.3.2. Introduction to clustering algorithm software

Since K-means cluster analysis requires multiple iterations, software is used for calculation. The data mining software is SPSS modeler, a comprehensive data mining software. Its advantage is that the operation is relatively simple, the data processing interface is direct, and the data can be analyzed by the flow method to obtain the result. It can often be found that there are many places in the traditional forecasting process that cannot be analyzed in detail, and analogy estimates can only be made based on experience and past cases. As a result, it is often distorted when estimating support capabilities. And because of the variety of influencing factors in the actual production process, there may be a lack of traditional analysis. At this point, using relevant data processing software to aid in data processing can reduce the workload and consider more relevant factors. The following uses SPSS Modeler to perform data mining on an airport. The SPSS Modeler is set up as follows: Log-likelihood estimation is used for clustering in two-step clustering. After clustering, you can try to analyze the association rules with the priori method, and further explore the relevant rules after clustering.

### 3.3.3. Calculation process

The purpose of two-step clustering is to identify key influencing factors. Take the data of an airport as an example to perform two-step clustering. The flow chart is shown in the figure below. The specific settings in the type are as follows. In the format, you need to pay attention to the type of time unit, and the rest can be used by default.
The results of running two-step clustering are as follows:

**Figure 5. Clustering results of airport support data**

| Cluster | Voyage time | Plane model |
|---------|-------------|-------------|
| Clustering-1 | 1 hour (72.7%) | A319 (45.5%) |
| Clustering-2 | 1.5 hours (40.0%) | A320 (50.0%) |
| Clustering-3 | 2 hours (66.7%) | E145 (66.7%) |
| Clustering-4 | 2.5 hours (50.0%) | B738 (100.0%) |
| Clustering-5 | 3 hours (100.0%) | B737 (100.0%) |

After clustering, you can visually see the clusters that are closely related. By clustering the results, it can be seen that the cluster with higher importance is a short-haul flight with the voyage time of 1 hour and the A319 model. After the clustering, an association analysis of prion is attempted. After the confidence value is set to less than 50%, the association rule cannot be generated, indicating that the association rule does not exist in this data table, and only the clustering rule exists.

From the results of the operation, it can be known that the airport needs to pay more attention to the short-haul flights with a flight time of 1 hour and the combined flight of the Airbus A319 during the peak season of the summer vacation, and to improve the apron support capability.

4. Conclusion

After cluster analysis of the peak period data of the airport, we found a cluster of high importance. In the ground operation of the airport, it is necessary to make early plans for this type of flight, improve
the punctuality rate of the release and the passenger satisfaction with the airport work. It should be noted that the results of K-means cluster analysis for flights at different airports will be different. The result of a combination of a one-hour short-haul flight and the Airbus A319 aircraft as a key factor applies only to the Airport. However, the cluster analysis method can provide a useful idea for the information processing and mining of the comprehensive support capability of different airports.

References
[1] Liao Dan, Huang Baojun. Simulation and evaluation based on AirTOp airport comprehensive support capability. Aeronautical Computing Technology, Vol. 1 (2017) No. 47, p.77 - 80.
[2] Xu Ruishi. Evaluation and optimization of ground support capacity at busy airports. Nanjing: Nanjing University of Aeronautics and Astronautics, (2014).
[3] He Xiaojun. Multivariate statistical analysis. Beijing: China Renmin University Press, (2015).
[4] Zhang Wenyu, Wang Xiuxiu, Ren Lu and so on. Application of improved principal component clustering analysis in education informationization, Vol. 7 (2017) No. 36, p.