Study on "three places and four fields" airline resource cooperative allocation based on decision tree algorithm

Dingfa Luo*1
Xinjiang Airport (Group) Co., Ltd.; Urumqi, Xinjiang, 830016, China

Abstract: In view of the disharmony development status of Beijing Tianjin Hebei region, the paper puts forward a route optimization Principle of cooperative allocation of route resources of airport group from the angle of flight punctuality, airlines share and function orientation. An optimal model of airline resources cooperative allocation is established to maximize the overall operational benefit of regional airports and the modified decision tree C4.5 algorithm was applied to solve the problem. Since this algorithm combined with constraint condition of model as pruning standard, the efficiency of the algorithm was greatly improved. Finally, the problem find the optimal solution could be found by comparing the profit and loss value of the program. The analysis example showed that the overall benefit of multi-airport system increased by 7.03%. This model was useful to optimize airline network and conformed to the actual situation, and the research results can provide a reference for civil aviation authorities in planning the airline network of multi-airport system.

1 Introduction

In December 2014, the Strategic Cooperation Framework Agreement on the Coordinated Development of Beijing-Tianjin-Hebei Airports was signed in Beijing, marking the implementation of the Beijing-Tianjin-Hebei coordinated development strategy under the new normal and taking the lead in realizing coordinated development in the airport field. In 2019, Beijing Daxing International Airport, the comprehensive transportation hub of the Beijing-Tianjin-Hebei region, officially opened to traffic, and since then, the layout of the Beijing-Tianjin-Hebei region's "three places and four airports" air route network has been formed. However, the development of Beijing-Tianjin-Hebei airport cluster is not harmonious at this stage. Beijing Capital Airport has obtained a large number of route resources, but the route network structure is not perfect, especially the frequency of North American international routes is lower than Incheon International Airport and Narita International Airport; The construction of Daxing Airport further changes the functional orientation of Beijing two airports; In addition, the passenger flow of Tianjin Airport and Shijiazhuang airport is insufficient, and the utilization rate of airport infrastructure is low, which leads to the waste of time slot and route resources, and affects the improvement of the overall efficiency of the airport group. Therefore, how to consider the different interests of different airports in the airport group to rationally allocate route resources and realize the "three places and four places" coordinated development of Beijing-Tianjin-Hebei is an urgent problem to be solved at present.

At present, the development level of domestic airport groups needs to be improved, and the work on airline network layout and management mode innovation of airport groups needs to be further strengthened. In a multi-airport area, passengers must decide which airport and which airline to provide travel services. Therefore, Pels et al.[1] established a discrete choice model of airport-airline joint selection by using nested logit model, considering passengers' preferences for airports and airlines, which provided a method for analyzing airport competition in a multi-airport environment. Bai Mingguo[2] conducted a comprehensive study on airline route network optimization, conducted a quantitative comparative study on fully connected route network and hub route network, and proposed a three-stage method for route network design considering flight planning in route network construction. Aiming at the optimization of internal and external route network construction of regional multi-airport system, Wang Lu et al.[3] adopted 0-1 integer programming model to realize the overall allocation of regional route resources, which not only met the demand of route traffic, but also relieved the pressure of hub airports, thus improving the operational efficiency of the entire regional multi-airport system. Zhang Jie[4] used the multi-attribute decision-making method to evaluate the compatibility of the current routes and functional positioning of the Beijing-Tianjin-Hebei airport Group, so as to find out the routes to be optimized and relocate the flights on the routes to be optimized to each airport in the

1 Email: 252323341@qq.com
airport group. Song Nie et al.\cite{5} established a regional multi-airport system route network optimization model to solve the problem of serious homogeneity among airports in the regional multi-airport system, and adopted the method of combining search tabu algorithm and shortest circuit algorithm to solve the problem, which reduced the homogeneity of routes. However, their understanding of airport groups was flawed, and the credibility of the selected research objects was poor. The universality of the model needs to be further proved. In addition, decision tree C4.5 algorithm, as one of the ten classic algorithms in data mining, has been widely used in machine learning, knowledge discovery and other fields. The C4.5 algorithm (Classification 4.5) proposed by Quinlan\cite{6} is a classic algorithm for generating decision trees and an extension and optimization of ID3 algorithm. It mainly improves ID3 algorithm in the following aspects: (1) Split attribute is selected by information gain rate; (2) The continuous attributes are discretized; (3) Pruning operation is carried out after constructing the decision tree; (4) Able to process training data with missing attribute values. However, the C4.5 algorithm is only suitable for data sets that can reside in memory, and the program cannot run when the training set is too large to fit in memory. Cheng Long et al.\cite{7} proposed an improved incremental learning method to address the shortcoming of C4.5 algorithm that is not capable of incremental learning.

In this paper, a cooperative allocation optimization model of airport group route resources aiming at maximizing the overall benefit of the multi-airport system is established, and the improved decision tree C4.5 algorithm is adopted to solve it. The obtained route resource allocation scheme is closer to the actual operation situation, and the operation efficiency of the regional multi-airport system is greatly improved.

2 Principle of cooperative allocation of airport group route resources

This paper focuses on solving the problem of how to coordinate the allocation of "three places and four airports" route resources, allocating the traffic resources of Capital Airport to Daxing Airport, Tianjin Airport and Shijiazhuang Airport, alleviating the operating pressure of Capital Airport, and guiding the formation of a more reasonable regional route network structure. Taking this as the starting point, this paper considers the following three parameters to carry out the research on the cooperative allocation of route resources in the Beijing-Tianjin-Hebei airport group:

(1) Flight punctuality
Flight punctuality rate is one of the important indexes used to judge the operation efficiency of airport. At present, the capacity of Capital International Airport is saturated, the time-slot resources are tight, and the route congestion leads to the increasing cost of flight delay; On the contrary, Daxing Airport, Tianjin Airport and Shijiazhuang Airport, by contrast, the passenger and cargo turnover does not match the airport construction scale, the utilization rate of route resources is low, and the airport capacity is excessive.

(2) Airline share
CAAC issued various preferential policies to encourage airlines to set up branch offices, base airlines and low-cost airlines in Tianjin Airport and Shijiazhuang Airport; At the same time, major airlines are increasing their share of capacity at their main operating airports.

(3) Function orientation
The homogenization competition among domestic regional airports is serious, and it is necessary to take differentiated development routes and distinguish functional positioning. For example, the main operation focuses on international or domestic, transportation or general navigation, passenger transport or freight, hub or branch, etc., and give full play to their own advantages, layout a more reasonable route network structure based on function optimization, serving the region and spreading out.

3 Airport group route network optimization model

3.1 Model assumption
(1) To ensure the applicability of the model, this paper makes the following assumptions:

(2) Assume that the highest administrative units of each airport in the airport group are consistent, there are fewer obstacles to high-level policy issuance and implementation, and the airports can cooperate smoothly.

(3) Assume that the ground communication between airports in the airport group system is convenient, diverse and accessible, and passengers can realize free transit between airports in the airport group.

(4) Within the airport cluster, large hub airports have saturated capacity and tight time slot resources, while small and medium-sized airports have relatively surplus capacity and low time slot utilization.

The same evaluation criteria shall be adopted for flight slots within the same time period.

3.2 Objective function

According to the principle of cooperative allocation of airport group route resources, the optimization model of cooperative allocation of airport group route resources is established and solved with the aim of maximizing the overall benefit of multi-airport system.

\[ F(i,j,u,v) = \max \sum_{i} \left( \sum_{j} \sum_{u} \sum_{v} Y_{ij}^v X_{ij}^v \right) \]

\[ Y_{ij}^v = \left( \alpha T_{ij}^v + \beta L_{ij}^v + \gamma L_{ij}^v \right) / \left( \alpha + \beta + \gamma \right) \]

In formula (1) (2), \( T_{ij}^v \) is the flight delay probability index, indicating the on-time rate of flights from airport i to airport j operated by Airline u at time v, \( A_{ij}^u \) refers to the market share of u airline at Airport i, representing the proportion of flights operated by u Airline at time v to all flights on the route from Airport i to Airport j at that time; \( L_{ij}^v \) refers to the degree of fit between the routes operated by u Airlines at the airport i at the moment v and the functional positioning of the airport i.

{\begin{align*}
L_{ij}^v \geq 0 & \text{, The route is in line with the function positioning of i Airport} \\
L_{ij}^v < 0 & \text{, Other}
\end{align*}}
represent the value of the route from Airport i to Airport j operated by u Airlines at time v, The weighted average value of the three evaluation indicators $T_{ij}^u$, $A_{ij}^u$ and $L_{ij}^u$, $\alpha \times \beta$ and $\gamma$ are adjustment coefficients: The dependent variable $x_{ij}^{uv} \in \{0,1\}$, This route does not exist or is cancelled, indicate s whether a flight of u Airlines chooses this route; The dependent variable $F(i,j,u,v)$ is the final solution of the objective function, $\forall i \in \mathcal{I}, \forall j \in \mathcal{J}, \forall u \in \mathcal{U}, \forall v \in \mathcal{V}$.

Set $I=[1,2,\ldots,l], J=[1,2,\ldots,m], U=[1,2,\ldots,n]$ are all integers. The set $V$ represents the set of time slots.

### 3.3 Constraint condition

1. **Peak hour capacity limit.** Airports have posted maximum operating capacity limits, so the number of inbound and outbound flights during peak hours needs to be taken into account.

$$\sum_{i \in \mathcal{I}} \sum_{u \in \mathcal{U}} x_{ij}^{uv} \leq W_{iv}$$  \hspace{1cm} (3)

2. **The degree of connectivity between airports.** The number of routes opened is one of the important indicators of the size of the airport, which can increase the choice of passengers, so that every airport will have at least one airport to fly with it.

$$\sum_{i \in \mathcal{I}} \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} x_{ij}^{uv} \geq 1$$  \hspace{1cm} (4)

3. **Uniqueness of the flight.** Any approach or departure flight can only choose to land or take off at one airport and within one time slot.

$$\sum_{i \in \mathcal{I}} x_{ij}^{uv} = 1$$  \hspace{1cm} (5)

4. **The airline and the flight are consistent at all times.** In order to improve the applicability of the model, it is assumed that the operating routes, destination airports and flight slots of the airlines in the airport group are fixed for a long time except the departure airports are undetermined. Therefore, if the profit and loss value of different take-off airports in the selected airport group is greater than the initial solution, the solution result of the model will be replaced with the new configuration result scheme; otherwise, the original configuration scheme will remain unchanged.

$$F(i,j,u,v) = \max \left\{ G(i,j,u,v), x_{ij}^{uv} \right\}$$  \hspace{1cm} (6)

In formula (3) - (6), $W_{iv}$ is the maximum operating capacity limit of i airport at moment v; $F(i,j,u,v)$ is a subset of all airport group route resource cooperative allocation schemes obtained by model optimization.

$F(i,j,u,v) = \max \left\{ x_{ij}^{uv} \mid i \in [1,2,\ldots,l], \forall v \in [1,2,\ldots,l] \right\}$ and i is an integer; $X_{ij}^{uv}$ is the set of routes to be optimized; that is, the initial route resource allocation scheme input by the model: $x_{ij}^{uv}$ is the on-time rate obtained from the initial route resource allocation scheme.

### 4 Model solving

In this paper, the improved decision tree C4.5 algorithm is applied to the cooperative allocation optimization problem of airline resources, and the specific process is shown in Figure 1 Decision tree C4.5 algorithm is the most mature concept learning method developed so far, and is the main technology for classification and prediction. It uses information gain rate to characterize the characteristic properties of each branch. Its basic principle is that the branch nodes represent the problem to be decided, different branches represent each configuration scheme, and probability values represent the possible situations of each configuration scheme. Profit and loss values of each scheme are calculated and compared to provide decision-making basis for decision makers. The cooperative allocation optimization problem of route resources studied in this paper is described as follows: decision tree nodes and branching schemes are taken as independent variables at the same time, the sum of all route evaluation values of each scheme is the profit and loss value, and the optimal allocation scheme is the decision tree branch with the maximum profit and loss value under the constraint conditions.

Step1: Samples were selected for data preprocessing.

Step2: Calculate the route valuation.

Step3: The construction of decision tree is divided into two steps: the generation of decision tree and the pruning of decision tree.

Establish decision tree. The decision tree construction method refers to taking each single attribute as a child node of a tree through attribute classification conditions, and then connecting each child node. In this paper, C4.5 decision tree algorithm is used to generate decision tree, which selects test attributes according to information gain rate. Information gain rate definition:

$$\text{GainRatio} = \frac{\text{Gain}(D,A)}{\text{SplitInfo}(D,A)}$$  \hspace{1cm} (7)

Information gain rate Normalizes information gain using the "split information" value. The split information is defined as follows:

$$\text{SplitInfo} = \sum_{j=1}^{k} \frac{D_j}{D} \log_2 \left( \frac{D_j}{D} \right)$$  \hspace{1cm} (8)

Where, A represents A variable with V values, $j \in \{1,2,\ldots,v\}$, and D represents the set of data where A takes the JTH value. Gain is information gain. The split information value represents the set of data where A takes the JTH value. Gain is information gain; The split information value SplitInfo represents the information generated by partitioning the training data set D into V partitions corresponding to v outputs of the attribute A test. By analogy, the information gain rate of each attribute is calculated, and the one with the largest information gain rate is taken as the root node of the decision tree and iteratively branched to form a decision tree.

Trim the branches. In this paper, based on the traditional pessimistic pruning method, the model constraints are transformed into pruning conditions to improve the C4.5 algorithm. Pruning is carried out in the process of generating decision trees, and the growth of trees can be stopped if the constraints are not met.

Step4: Calculate the profit and loss value of each scheme.

Step5: Get the optimal scheme.
5 Case verification

5.1 Sample selection and parameter design

In Beijing-Tianjin-Hebei "three to four games" as the research object, from the global real-time flight transportation web site (https://zh.flightaware.com/) for November 23, 2019, 8: The average delay rate of Capital Airport, Tianjin Airport and Shijiazhuang Airport is 61.55%, 70.62% and 73.66%, respectively. In this example $\alpha = \beta = \gamma = 1$.

5.2 Experimental results and analysis

This paper uses Matlab programming to implement the algorithm. The sample data set shown in Table 1 contains a total of 11 flights to be optimized. According to the selection of departure airport i, each flight has 3 route selection scheme branches and corresponding route evaluation value, and a total of 311 route configuration schemes, which are marked as YES or NO according to the profit and loss value of each scheme. The improved C4.5 algorithm is used to construct the decision tree and cut the NO branches.

<table>
<thead>
<tr>
<th>Number</th>
<th>j (Shenzhen Airport)</th>
<th>u (Xi'an Airport)</th>
<th>v (Chongqing Airport)</th>
<th>$Y_{ij}^{uv}$ (i= Capital Airport)</th>
<th>$Y_{ij}^{uv}$ (i= Tianjin Airport)</th>
<th>$Y_{ij}^{uv}$ (i= Shijiazhuang Airport)</th>
<th>SplitInfo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Shenzhen Airport</td>
<td>ZH</td>
<td>8</td>
<td>0.209121</td>
<td>0.270511</td>
<td>0.262533</td>
<td>0.605275</td>
</tr>
<tr>
<td>2</td>
<td>Shenzhen Airport</td>
<td>CA</td>
<td>8</td>
<td>0.329808</td>
<td>0.288710</td>
<td>0.276937</td>
<td>0.605209</td>
</tr>
<tr>
<td>3</td>
<td>Xi'an Airport</td>
<td>CZ</td>
<td>8</td>
<td>0.317922</td>
<td>0.268595</td>
<td>0.291340</td>
<td>0.605300</td>
</tr>
<tr>
<td>4</td>
<td>Xi'an Airport</td>
<td>MU</td>
<td>8</td>
<td>0.315621</td>
<td>0.255185</td>
<td>0.293398</td>
<td>0.605289</td>
</tr>
<tr>
<td>5</td>
<td>Chongqing Airport</td>
<td>NH</td>
<td>8</td>
<td>0.272357</td>
<td>0.242733</td>
<td>0.262533</td>
<td>0.606205</td>
</tr>
<tr>
<td>6</td>
<td>Chongqing Airport</td>
<td>G5</td>
<td>8</td>
<td>0.277864</td>
<td>0.255185</td>
<td>0.270764</td>
<td>0.606202</td>
</tr>
<tr>
<td>7</td>
<td>Haikou Airport</td>
<td>HU</td>
<td>8</td>
<td>0.277657</td>
<td>0.259016</td>
<td>0.266649</td>
<td>0.606247</td>
</tr>
<tr>
<td>8</td>
<td>Hefei Airport</td>
<td>CA</td>
<td>8</td>
<td>0.366792</td>
<td>0.288713</td>
<td>0.276937</td>
<td>0.601785</td>
</tr>
<tr>
<td>9</td>
<td>Fuzhou Airport</td>
<td>SC</td>
<td>8</td>
<td>0.223345</td>
<td>0.252318</td>
<td>0.270764</td>
<td>0.605236</td>
</tr>
<tr>
<td>10</td>
<td>Kunming Airport</td>
<td>GS</td>
<td>8</td>
<td>0.227848</td>
<td>0.375879</td>
<td>0.262533</td>
<td>0.596979</td>
</tr>
<tr>
<td>11</td>
<td>Hong Kong Airport</td>
<td>CA</td>
<td>8</td>
<td>0.725677</td>
<td>0.288714</td>
<td>0.276937</td>
<td>0.601412</td>
</tr>
</tbody>
</table>

Table 2 The comparison between the decision tree C4.5 algorithm and the modified decision tree C4.5 algorithm

<table>
<thead>
<tr>
<th></th>
<th>C4.5 Algorithm</th>
<th>Improved C4.5 Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaf node number</td>
<td>411</td>
<td>2985</td>
</tr>
<tr>
<td>Depth of tree</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Average value of leaf node sample proportion</td>
<td>0.00%</td>
<td>0.05%</td>
</tr>
</tbody>
</table>

As can be seen from the comparison of Table 2, the complexity of the decision tree of the improved C4.5 algorithm is smaller than that of the traditional C4.5 algorithm, in which the number of leaf nodes is reduced by 4191319, and the ratio of leaf nodes to samples is also larger than that of the decision tree established by the traditional C4.5 algorithm, thus greatly improving the efficiency and accuracy of the algorithm and making it more applicable to the problem studied.
As shown in Table 3, three flights have changed their departure airports, that is, three operational routes have been optimized. Among them, the Beijing-Shenzhen route operated by Shenzhen Airlines has been adjusted to the Tianjin-Shenzhen route; The Beijing-Chongqing route operated by China Airlines was adjusted to the Hebei to Chongqing route; The Beijing-Fuzhou route operated by Shandong Airlines has been adjusted to the Hebei to Fuzhou route. At the same time, four flights departing from Beijing were adjusted to take off from Daxing Airport. After optimization, the maximum profit and loss value of each route is 3.793116, which increases by 7.03% compared with the maximum of 3.544012 before optimization. The experimental results show that the route resource cooperative allocation scheme predicted by the model is more suitable for actual operation requirements, can improve the route network structure of the Beijing-Tianjin-Hebei airport group, and improve the overall operation efficiency of the regional airport group.

6 Summary

In view of the disharmonious development of “three places and four airports” in the Beijing-Tianjin-Hebei region, this paper proposes the principle of cooperative allocation of route resources of airport groups from multiple perspectives, establishes an optimization model of cooperative allocation of route resources aiming at maximizing the overall operating benefits of regional airport groups, and applies the improved C4.5 algorithm to the optimization of cooperative allocation of route resources. The algorithm combines the constraints of the model and trims the branches in the process of tree building, thus greatly improving the efficiency and accuracy of the algorithm. Finally, the optimal solution is found by comparing the profit and loss values of each scheme. The model and algorithm provide data analysis and optimization means for the study of airport group route resource cooperative allocation, and provide reference for finding the best airport group route resource cooperative allocation scheme. The next step is to optimize the reasonable allocation of airspace resources in the airport group from the Angle of air traffic control operation.

Reference


