Robust modeling for fleet assignment problem based on GA-SVR forecast

A J Su1, W D Yang2*, C Zhang and M X Kong

College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China

*Corresponding author’s e-mail: ywendong@nuaa.edu.cn

Abstract. Fleet assignment is the essential step of overall airline flight scheduling process, and the quality of assignment strategy directly affects economy and safety of air transportation. Market demand forecast is the premise of fleet assignment, and accurate forecast is an important guarantee to reduce passenger overflow and improve aircraft utilization rate. In order to carry out scientific fleet assignment, this paper studies from two aspects: Firstly, support vector machine regression (SVR) is used to forecast flight passenger flow and solve the problem of undeterminable parameters, we present a GA-SVR model with genetic algorithm for parameter optimization. Secondly, from the perspective of flight recovery efficiency, this paper incorporates the concept of fleet robustness, and establishes the robust model for fleet assignment with dual-objective, which maximizes the flight operating profit and minimizes the number of aircraft type in busy airports. Finally, flight network of an airline is analysed to verify the validity of the model and algorithm. It shows that: the MSE mean value of GA-SVR prediction results is between 0.0103 and -0.0031, which is relatively accurate. And the fleet assignment model can significantly improve robustness (16.7%) at the expense of less profit (4.2%).

1. Introduction
Fleet assignment is a process matching suitable aircraft types with various routes and flights to maximize the company’s revenue, which is also the basis of aircraft route planning and crew scheduling. It is based on the existing airline network and fleet resources of airline company and Considers factors such as fleet size, cabin capacity, operating cost and potential benefits [1]. Particularly, market size and characteristics is the important factor influencing the efficiency of fleet assignment, the consistency of aircraft size and market demand plays a critical role in creating higher economic benefits [2].

Compared with other modes of transportation, the existing research on the market demand of air passengers is not thorough yet. The econometric model [3-5] is mainly used to predict air passenger flow through analysing the influencing factors, but the factors are difficult to be determined accurately and more subjective. Conversely, machine learning algorithm focuses on analysing the law of data itself and is more objective. Among them, Support Vector Regression (SVR) is a more accurate method, the deficiency is that its results depend on the selection of kernel function parameters. In this regard, this paper proposes a GA-SVR prediction method, which uses Genetic Algorithm (GA) to improve the quality and efficiency of kernel function parameter optimization. For fleet assignment, A robust assignment strategy is conducive to the rapid recovery of irregular flights, which has great significance to the production and operation of airlines and the enhancement of passenger satisfaction.
Existing literatures mostly analyse using basic fleet assignment model [6-7]. There are few relevant studies on the robustness of fleet assignment, and the robustness index is relatively single [8], which has no good reference value for the adjustment of irregular flights recovery. Therefore, this paper improves the prediction method of market demand based on the SVR, and establishes the fleet assignment model with dual-objective: revenue and robustness. So as to better solve the problem of matching between demand and aircraft size.

2. Forecast of airline market demand based on GA-SVR

The airline market demand is a kind of time series, and its value is affected by many factors such as economy, weather, season, holiday, etc. Therefore, the forecast of demand is a complex nonlinear problem. In view of the above characteristics, GA-SVR method is adopted in this paper. In essence, VSR projects a given airline market demand sample set into high-dimensional space through nonlinear mapping, thus transforming nonlinear problems in original space into linear problems in feature space and reducing model complexity [9]. GA is used to optimize the kernel parameters of SVR, which has advantage of good robustness and universality. The specific algorithm flow is shown in the figure below.

\[
MSE = \frac{1}{n} \sum_{i} (y_i - \hat{y}_i)^2
\]  

(1)

In the equation, \(y_i\) is the \(i\)th observation, \(\hat{y}_i\) is the \(i\)th forecast and \(n\) is the quantity of sample. Cross Validation (CV) is used as parameter optimization criteria.

Step1: Selecting. The randomly initialized parameters \((c_0, g_0)\) are brought into the SVR model, the training sample data is input for model training, and the fitness value (mean square error) of sample is calculated according to the equation below. The highly adaptable individuals in the current population were selected to inherit the effective genes to the next generation by means of roulette.

Figure 1. GA-SVR algorithm flow char.

Step2: Crossing. The individuals in the population are randomly paired, the locations of the crossing points are randomly set by the single point intersection method, and some genes between the two individual chromosomes are exchanged by the crossover probability \(P_c\).

Step3: Mutation. Some genes of an individual are changed according to the probability of mutation probability \(P_m\), to adjust the condition that the fitness cannot reach the optimum after the crossing.
3. Robust model assignment model

During the flight operation, it is often affected by weather, flow control and other interference, which can lead to flight delay, and the delay of a single flight will have ripple effects, generating large-scale irregular flights in the hub airport. Therefore, it is of great practical significance to take external disturbances into consideration when planning, and make robust fleet assignment plan. Considering that busy airports are prone to delays, and exchange between aircrafts of the same type is the main and the least loss strategy for flight recovery, this paper proposes a two-objective assignment model with the least type of aircraft in busy airports and the largest profit.

3.1 Profit of fleet assignment

Flight profit is obtained by deducting the cost from the revenue:

\[ E_{ij} = P_{ij} - C_{ij} \]  

\[ P_{ij} \] refers to the revenue of flight \( i \) executed by type \( j \). It is calculated by taking the larger value of passenger number \( pax_i \) and seat number \( s_i \) multiplied by the average passenger ticket price \( price_i \):

\[ P_{ij} = \max\{pax_i, s_i\} \times price_i. \]

\[ C_{ij} \] refers to the total cost, including the aircraft operation cost \( OC_{ij} \) and passenger overflow cost \( SC_{ij} \). \( OC_{ij} = a_j + b_j T_i s_j \). \( a_j \) and \( b_j \) are the fixed cost and variable cost coefficients of type \( j \), \( T_i \) is the chock time of flight \( i \), and \( s_j \) is the seat number of model \( j \). \( SC_{ij} = (\min\{pax_i, s_i\} - s_i) \times price_i \), which means that take the part of \( pax_i \) that exceeds \( s_i \) multiplied by the average passenger fare \( price_i \).

3.2 Fleet assignment model establishment

A space-time network model \( G(N, A) \) is established. For each aircraft type \( j \), node \( N \) represents the airport-time two-dimensional point, and arc \( A \) represents the flight [10]. The following is the specific model parameters: \( K \) denotes the aircraft type set, \( j \in K \); \( n_j \) is the aircraft number, \( s_j \) is the seat number; \( F \) stands for flight set, \( i \in F \); \( O \) stands for the set of stop-over flights, \( O \subseteq F \), \( o \in O \). Most domestic stop-over flights only have one stop, which include two flight segments \( o_1 \) and \( o_2 \). \( A \) denotes the airport set, \( a \in A \); \( m_a \) refers to the type number of aircraft taking off and landing at airport \( a \); \( H \) stands for busy airport set, \( H \subseteq A \); \( M \) is the space-time network node, \( (ak) \in M \), \( (ak') \) is the previous time node right next to \( (ak) \), \( (ak') \) is the last time node of airport \( a \); \( e_{(ak)} \) is a constant. When flight \( i \) lands at the node \( k \) of airport \( a \), it equals 1, when it takes off, it equals -1. When neither, it is 0; \( f_{oa} \) is a type 0-1 indicator operator. It is equal to 1 when flight \( i \) takes off or lands at airport \( a \), otherwise it is 0. \( G_{(ak)j} \) is an integer variable, representing the aircraft number of type \( j \) at the node \( k \) of airport \( a \); \( x_{ij} \) is 0-1 decision variable, if flight \( i \) is executed by type \( j \), it is 1, otherwise, it is 0; \( z_{aj} \) is 0-1 decision variable, which is equal to 1 when type \( j \) takes off or lands on airport \( a \), otherwise 0. The model is established as follows:

\[ \min \quad w = \sum_{a \in H} m_a \]  

\[ \max \quad z = \sum_{i \in F} \sum_{j \in K} (P_{ij} - C_{ij}) x_{ij} \]  

\[ \text{s.t.} \quad \sum_{j \in K} x_{ij} = 1, \forall i \in F \]  

\[ G_{(ak')j} + \sum_{i \in F} e_{(ak)j} x_{ij} = G_{(ak)j}, \forall (ak) \in M, j \in K \]  

where \( E_{ij} \) is the fitness of new individuals generated by crossing and mutation is calculated. And then the new generation group is formed together with the parent generation.

Step 5: Check whether the maximum number of population or maximum number of iterations which are pre-set can be reached. If so, terminate the operation, and get the optimal parameter \((c, g)\) to predict the test set; otherwise, go back to step 2.
\[
\sum_{ac} G_{(ac)} \leq n_j, \forall j \in K, (ab) \in M
\] (7)

\[
f_{a_i} x_{ij} \leq z_{ij}, \forall i \in F, j \in K, a \in H
\] (8)

\[
\sum_{j \in K} z_{ij} \leq m_a, \forall a \in H
\] (9)

\[
x_{aj} - x_{aj} = 0, \forall j \in K, o \in O
\] (10)

\[
x_{ij} = 0,1, \forall i \in F, j \in K
\] (11)

\[
G_{(ab)} \in z^*, \forall (ab) \in M, j \in K
\] (12)

Equation (3) is the first objective function, with the minimum type number of takeoff and landing in busy airports, and equation (4) is the second, with the maximum profit; Equation (5) is flight coverage constraint; Equation (6) refers to the flow balance. It means that in an airport, the number of planes at a time node is equal to the one at the next time node after the change of takeoff and landing. Equation (7) represents the conservation of aircraft quantity; Equation (8) means that only when the aircraft of type \(j\) takes off or lands at airport \(a\), the flight \(i\) on airport \(a\) can be executed by type \(j\). Equation (9) means that the type number of aircraft taking off or landing at airport \(a\) is no more than \(m_a\). Equation (10) specifies that aircrafts assigned for the two sections of a stop-over flight must be of the same type.

This model belongs to dual-objective programming, which is very difficult to solve. Therefore, this paper transforms it into a single-objective programming. Firstly, the single-objective model of profit maximization can be obtained by removing the equation (3), and the result of the model is the maximum profit value \(E_m\). Then, according that proportion of total profit reduction after increasing the robustness is specified to not exceed \(r\), the objective equation (4) is turned into the constraint condition below.

\[
\sum_{i \in F} \sum_{j \in K} (P_{ij} - C_{ij}) x_{ij} \geq (1 - r)E_m
\] (13)

4. The example analysis

In order to verify the rationality of the model and algorithm, flight network of a large domestic airline company is used for analysis. The case involves a total of 12 aircraft of 3 types, executes 56 flight segments on 21 routes, of which 20 are stop-over flights. All flight numbers, arrival and departure time, take-off and landing airports are read from flight schedules, and the average passenger ticket price is calculated by dividing the annual total revenue of the airline market by the actual passenger volume. The aircraft number, seats and costs of all types are shown in table 1, in which the cost data of different types is obtained by SPSS regression analysis. Market demand forecast uses passenger flow data of 96 months ranging from 2010 to 2017 as the sample.

| Type | Number | Seat | Fixed cost  | Variable cost coefficient |
|------|--------|------|-------------|--------------------------|
| A319 | 6      | 122  | 23666.97    | 3.02                     |
| A321 | 2      | 177  | 32583.89    | 2.10                     |
| A325 | 4      | 186  | 36171.47    | 1.92                     |

4.1 Market demand forecast

The sample data from 2010 to 2016 is taken as the training set and 2017 as the test set. The parameters of the algorithm are set as follows:

\[
\sum_{a,c} G_{(a,c)} \leq n_j, \forall j \in K, (ab) \in M
\] (7)

\[
f_{a_i} x_{ij} \leq z_{ij}, \forall i \in F, j \in K, a \in H
\] (8)

\[
\sum_{j \in K} z_{ij} \leq m_a, \forall a \in H
\] (9)

\[
x_{aj} - x_{aj} = 0, \forall j \in K, o \in O
\] (10)

\[
x_{ij} = 0,1, \forall i \in F, j \in K
\] (11)

\[
G_{(ab)} \in z^*, \forall (ab) \in M, j \in K
\] (12)
Table 2. GA-SVR algorithm parameters.

| Algorithm | Parameter                  | Value             |
|-----------|----------------------------|-------------------|
| GA        | Maximum population         | 20                |
| GA        | Maximal evolution iterations| 200               |
| GA        | Crossing probability       | 0.9               |
| GA        | Mutation probability       | 0.09              |
| SVR       | Range of penalty coefficient | $2^{-2}$ — $2^{10}$ |
| SVR       | Range of kernel parameter  | $2^{-15}$ — $2^{15}$ |

GA-SVR is used for fitting and prediction, and compared with the simple SVR algorithm. 5 flight segments are taken as examples to analyze. Table 3 is the result of VSR parameter optimization, and the optimal parameters obtained by the two methods are significantly different. The results of training set fitting and test set prediction in Chengdu-Wuxi are drawn into broken line graphs, as shown in figure 2 and 3, indicating that the GA-SVR results are relatively accurate.

Table 3. The VSR parameter optimization results of the two algorithms.

| Flight segment       | Algorithm | Parameter c | Parameter g |
|----------------------|-----------|-------------|-------------|
| Chengdu--Wuxi        | GA-SVR    | 54.98       | 2937.01     |
| Guangzhou--Wuxi      | SVR       | 4.00        | 1552.09     |
| Nanjing--Chengdu     | GA-SVR    | 32.18       | 2031.07     |
| Nanjing--Shenzhen    | SVR       | 0.44        | 294.07      |
| Nanjing--Xiamen      | GA-SVR    | 2.43        | 402.69      |
|                      | SVR       | 1.32        | 512.00      |
|                      | GA-SVR    | 0.28        | 3978.45     |
|                      | SVR       | 0.25        | 1552.09     |
|                      | GA-SVR    | 125.84      | 2994.17     |
|                      | SVR       | 4.00        | 97.01       |

Figure 2. Fitting results of training set.  
Figure 3. Prediction results of test set.

By analyzing prediction accuracy of the model quantificationally, mean square error (MSE) is compared as shown in table 4. MSE reflects the deviation between the predicted results and the test sample values. The smaller the deviation, the more accurate the prediction. According to the table, the MSE values of GA-SVR are all low in the 12-month prediction of 5 routes, with the mean value between 0.0103 and -0.0031, generally less than that of SVR. This proves again that GA-SVR has a good prediction ability, and the optimization of parameters is better than simple SVR.

Table 4. MSE of the test set prediction results.

| Month | Chengdu-Wuxi | Guangzhou-Wuxi | Nanjing--Chengdu | Nanjing-Shenzhen | Nanjing-Xiamen |
|-------|--------------|----------------|------------------|------------------|----------------|
|       | GA-SVR       | SVR            | GA-SVR           | SVR              | GA-SVR         |
| 1     | 0.0091       | 0.0090         | -0.0091          | 0.0075           | -0.0541        |
| 2     | -0.0076      | -0.0254        | 0.0091           | 0.0451           | 0.1996         |
| 3     | 0.0105       | 0.2374         | -0.0083          | -0.0576          | 0.1729         |
4.2 Robust fleet assignment

According to the characteristics of airline routes and operation bases, the busy airports are set as Nanjing and Shanghai. This paper assumes that the profit reduction should not be higher than 5% after considering the robustness. Firstly, the goal is to maximize profit, and the maximum profit is 177,4697 ¥, with robustness 6. It means that there are 3 types of take-off and landing aircraft in both two busy airports, which is not conducive to the implementation of the replacement strategy. Then, the robust factor is added, solving the dual objective programming can finally obtain the assignment scheme shown in Table 5. According to the results, the profit is 1700160 ¥, and the robustness is 5. The operating profit decreased by 4.2% after taking into account the robustness, but the robustness increased by significant 16.7%, which can better deal with irregular flights.

Compared with the actual manual assignment plan using in the airlines now and the robust fleet assignment plan of this paper, the number of aircrafts needed for all flights was reduced by 1, and the total aircraft utilization rate increased from 617.92h to 674.09h, which increased by 9.1%, indicating that the model in this paper can make better use of aircraft resources. The number of flights executed by A319 has decreased by 4 compared with the actual number, and the type number and the utilization rate have decreased, while A321 is just the opposite. It indicates that A321 is more economical than A319. Although the number of flights executed by A325 remains unchanged, both the aircraft type and the aircraft utilization rate have increased, its economy is also better. However, it has the largest number of seats, which leading to high cost. Therefore, when the existing passenger volume cannot reach the high passenger rate, it should choose small aircraft to execute, thus saving an aircraft.

Table 5. The results of the robust fleet assignment compared with the actual manual assignment.

| Type   | Actual manual assignment | Robust fleet assignment |
|--------|--------------------------|-------------------------|
|        | Aircraft number | Flight number | Utilization rate of type /h | Utilization rate of aircraft/h | Aircraft number | Flight number | Utilization rate of type /h | Utilization rate of aircraft/h |
| A319   | 6            | 32            | 3795                       | 632.50                       | 6              | 28            | 3515                       | 585.83                       |
| A321   | 2            | 8             | 1230                       | 615.00                       | 2              | 12            | 1450                       | 725.00                       |
| A325   | 4            | 16            | 2390                       | 597.50                       | 3              | 16            | 2450                       | 816.67                       |
| Total  | 12           | 56            | 7415                       | 617.92                       | 13             | 56            | 7415                       | 674.09                       |

5. Conclusion

This paper studies the two important links with closely relation in airline flight scheduling - market demand forecast and fleet assignment. Taking into account the complex nonlinear variation law of air
passengers and the problem of limited samples, SVR is used to predict. The SVR parameters are optimized by GA, which greatly improves the efficiency and accuracy. In terms of aircraft fleet assignment, In order to relieve the serious delay in busy hub airports, a dual-objective assignment model with the least number of types in busy airports and the largest profit is established. It can increase the opportunity to exchange aircrafts between the same types when flight recovery, and improve the robustness of assignment. The model and algorithm have important practical significance for alleviating flight delay, and can provide good theoretical basis and effective decision support for airline operation.

Acknowledgment
1Fund project: Nanjing university of aeronautics and astronautics graduate innovation base (laboratory) open fund (kfjj20170714, kfjj20170702)
2Fund project: Youth science and technology innovation fund (NS2016063)
Fund project: Fundamental research funds for the central universities (NZ2016109)

References
[1] Zhu X H, Zhu J F and Gao Q 2012 Science Technology and Engineering, 12 1329-33
[2] Wu D H and Xia H S 2014 Journal of Transportation Systems Engineering and Information Technology. 14 109-37
[3] Peng T, Zhang Y P and Hao S Q 2016 Journal of Transport Information and Safety. 34 37-44
[4] Preez J D and Witt S F 2003 International Journal of Forecasting. 19 435-51
[5] Yao T and Tao J 2015 Computer Technology And Development. 25 147-51
[6] Zhu B W and Tang F Y 2015 China Management Informationization. 18 135-36
[7] Tsai M W, Hong T P and Lin W T 2015 Mathematical Problems in Engineering. 1-12
[8] Zhu X H, Zhu J F and Gao Q 2011 Forecasting. 30 71-4
[9] Chen R, Liang C Y and Lu W X 2014 Systems Engineering — Theory & Practice. 34 1290-6
[10] Zhe L and Wanpracha Art C 2013 Transportation Science. 47 493-507