Identifying Similar Operation Scenes for Busy Area Sector Dynamic Management

HU Minghua[1*], *ZHANG Xuan*[1], *YUAN Ligang*[1], *CHEN Haiyan*[2], *GE Jiaming*[2]

1. College of Civil Aviation/College of Flight, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P. R. China;
2. College of Computer Science and Technology/College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P. R. China

(Received 10 June 2020; revised 12 July 2020; accepted 30 July 2020)

Abstract: Air traffic controllers face challenging initiatives due to uncertainty in air traffic. One way to support their initiatives is to identify similar operation scenes. Based on the operation characteristics of typical busy area control airspace, an complexity measurement indicator system is established. We find that operation in area sector is characterized by aggregation and continuity, and that dimensionality and information redundancy reduction are feasible for dynamic operation data base on principle components. Using principle components, discrete features and time series features are constructed. Based on Gaussian kernel function, Euclidean distance and dynamic time warping (DTW) are used to measure the similarity of the features. Then the matrices of similarity are input in Spectral Clustering. The clustering results show that similar scenes of trend are not ideal and similar scenes of modes are good base on the indicator system. Finally, actual vertical operation decisions for area sector and results of identification are compared, which are visualized by metric multidimensional scaling (MDS) plots. We find that identification results can well reflect the operation at peak hours, but controllers make different decisions under the similar conditions before dawn. The compliance rate of busy operation mode and division decisions at peak hours is 96.7%. The results also show subjectivity of actual operation and objectivity of identification. In most scenes, we observe that similar air traffic activities provide regularity for initiatives, validating the potential of this approach for initiatives and other artificial intelligence support.

Key words: air traffic; similar scenes; unsupervised clustering; dynamic operation; time series; similarity measure

CLC number: V355 **Document code:** A **Article ID:** 1005-1120(2020)04-0615-15

0 Introduction

Controlled airspace contains area control airspace, approach control airspace and tower control airspace, where the area control airspace connects different parts of approach control airspace and undertakes the navigation tasks of flights from different areas of approach control airspace. It is the main space of the controlled airspace that affects the safety and efficiency of flights. Identifying similar scenes is to measure the similarity of the time slice or time sequence with the operation feature in the airspace, and to identify the scenes with similar operation mode or operation trend. Through identification of similar operation scenes in the area control airspace, the historical operation situation can be effectively summarized[1]. Referring to the analysis of historical operation, making plan based on the future operation prediction is conducive to reduce the pressure of strategies formulation in the tactical stage, and to improve the operation efficiency of airspace. Also, it is the application foundation of artificial intelligence in control operation. For example, the recognition of meteorological images and radio-

*Corresponding author, E-mail address: minghuahu@nuaa.edu.cn.

How to cite this article: HU Minghua, ZHANG Xuan, YUAN Ligang, et al. Identifying similar operation scenes for busy area sector dynamic management[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2020, 37(4): 615-629.
http://dx.doi.org/10.16356/j.1005-1120.2020.04.012

telephony communications based on similar operation scenes is of great significance to the actual operation.

Most scholars in dynamic sector management focus on optimization methods, paying more attention to the sector design issues. Chen et al.[2] abstracted the dynamic sector problem as a graph partition problem using an improved genetic algorithm and gave the result of sector planning with sector workload balance as constraint. Serhan et al.[3] presented robust optimization and conditional value-at-risk (CVaR) approaches to mitigate the effect of weather forecast uncertainty by dynamically reconfiguring the terminal airspace. Tang et al.[4] used multi-objective optimization algorithm to give the results with maximum similarity and minimum load deviation. Yousefi[5] roughly divided the national airspace into three layers: Low-altitude, mid-altitude and high-altitude, and combined them in each layer using an optimal theoretical clustering algorithm. The obtained results are helpful to provide theoretical support for sector planning on the sector design level, but if the sector is already in operation, the control units prefer to analyze the current operation modes and deficiencies, and similar operation scenes identification can better satisfy their needs.

For identifying similar scenes in the air traffic management operation, researchers have done many studies on feature selection, feature extraction and similarity measurement. They also have made some achievements. In feature selection and feature extraction, the features used are mainly divided into two categories: Air traffic operation features, and meteorological features. Kuhn[6] selected the following methods based on expert knowledge: Distance from airport to heavy rainfall center, cross wind intensity of runway, number of scheduled arrival flights, etc. Christien et al.[7] took number of aircraft, type of aircraft, number of conflicts, number of peak hours, maximum performance of the aircraft and static sector indicators such as intersection, sector shape, sector size, etc. as input for classification. Andreeva-Mori et al.[8] chose airborne delay, ground delay and lost capacity to classify the mode of ground holding. Brinton et al.[9] transformed operational data such as number of aircraft, sector boundary proximity, flight time, heading, flight speed, etc, into dynamic densities as the clustering input. Gorripaty et al.[10] transformed all other relevant features into capacity features. Through the multidimensional scaling (MDS) visual analysis of the capacity flow data of four major airports in the United States, it was found that there was no natural class in capacity data and demand data, and that the dimensionality reduction of capacity data based on principal component analysis (PCA) was feasible. Arneson et al.[11] processed a large amount of weather and traffic data and ultimately converted it into WITI features to study the impact of convective weather on traffic. Grabbe et al.[12-13] used Jaccard′s coefficient to calculate the similarity of operation scenes and further used the measurement results to be the input for clustering and classification of similar scenes. Liu et al.[14] learned the measure matrix by defining similar and dissimilar scenes, using semi-supervised methods, and found the closest similar day from the measure matrix to the given day. Vargo et al.[15] proposed a random field model considering the high-dimensional, continuous classification changes, dynamic and other features of air traffic management data. By using end-user input to improve the learning of similarity scores, stable comparison of similarities for air traffic management was achieved. Gorripaty et al.[16] attempted to use the random forest model to learn airport random data using airport actual throughput, demand, and weather data. Then they got a similarity matrix using capacity and delay, and concluded that Euclidean distance is the most appropriate measure of cumulative capacity. Asencio[17] clustered U. S. convective weather based on Euclidean distance using K-means method, and finally selected representative days that could reflect U.S. convective weather impact patterns.

Most current studies are based on the overall national airspace or airport level, and there is little research on similar scenes in sectors. Those that focused on the discrete similarity measurement of horizontal operating features ignored the most significant vertical features of airspace relative to the ground and the time series features of flight continuity, thus lacking the study on discrete and time series similarity measurement about vertical sector operation. Identification of similar scenes using horizontal operation data facilitates the analysis of sector macro-strategies. Through the study of similar scenes of operation modes and trends of airspace sectors, it is helpful to guide the formulation of initiatives in operation level and to improve the flight operation efficiency. However, methods of selecting features and identifying similar scenes need to be studied.

From the perspective of sector-level regulation operation, this paper establishes the indicator system of area sector operation considering the level element, as shown in Fig.1. The values are calculated based on the indicator system, and the correlation of the features and the historical operation situation are analyzed. Then similarity of the discrete and time series data consisting of extracted features is measured. Using similarity measurement results, two types of similar operation scenes, operation mode and operation trend, are identified. Finally, the results of sector similar scenes identification is discussed, and scenes support is provided for the formulation of sector operation initiatives.

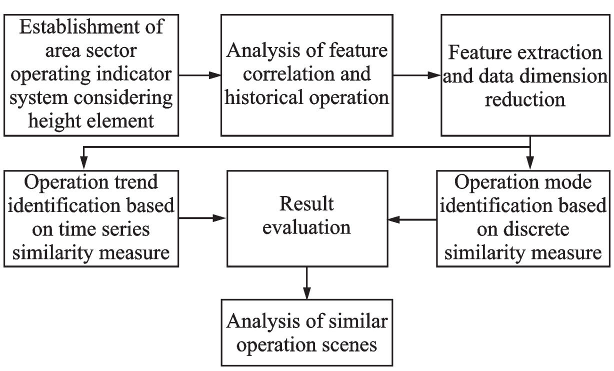

Fig.1 General framework

1 Analysis of Flight Operation Features in Area Sector

Before identifying the similar operation scenes of area sectors, it is necessary to establish an operation indicator system based on the operating features in area sectors, so as to calculate the value needed for similar scene identification.

Considering the possible redundancy of operating features, correlation analysis of operating features is needed to describe the sector operation. It determines whether feature extraction is performed, which will reduce the redundancy between features.

1.1 Definition of operating features

Area sectors are the main flying space of a flight. Different sectors have different number of waypoints, different airway structures, and different airspace areas. Weather is the environmental of a flight operation, which affects the overall operation of flights and the decision making of controllers. However, due to the higher altitude of the area sector, the area sector is less affected by the weather. Dynamic operating features are the most important features of flight situation, which represents the flight operation under influences of different space and weather.

For a particular sector, the static information of the sector can be assumed to be fixed. The dynamic operating features represent the flight situation under the influence of weather. There are inadequacies in the interpretation of sector operation by horizontal operation features in the vertical view, as shown in Fig.2. Under the operation in Fig.2(a), there are $(5+5+3)\times 3$ kinds of origin altitude positions, and each flight has three operating states if the three operation states are considered simply as level flight, descent and climb. Under the above condition, there are $(5+5+3)\times 3\times 3^3$ kinds of operation situations. Figs.2(b)—(d) shows three common operating situations. Therefore, this paper will select the dynamic operation features of flights

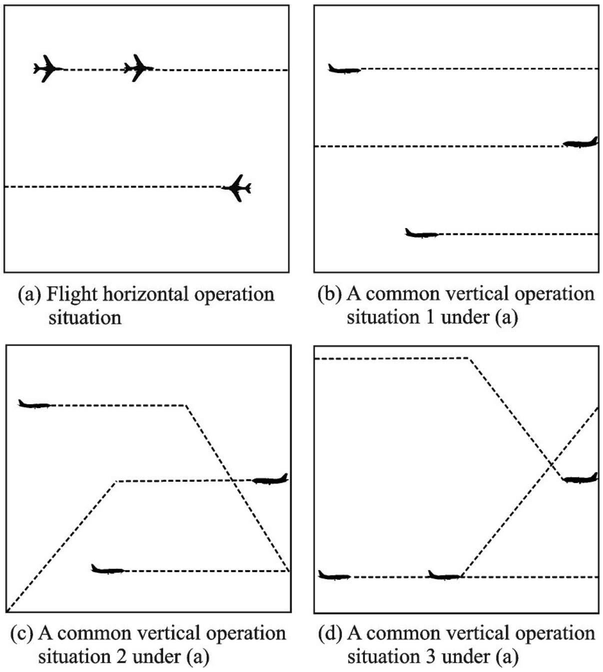

Fig.2 Typical flight operation considering level elements

based on level elements for identifying similar scenes.

In a certain operating environment, the operating situation will also show some regularity. For a particular sector, its static features can also be considered as a fixed amount. Therefore, this paper establishes the indicator system from the horizontal and vertical profiles.

Referring to Fig.2(a), operating features in terms of number of flights, convergence, distance/time, and average speed in horizontal profile dimension are given. For the vertical profile dimension, considering the three operation situations in Figs.2(b)—(d), the operating features from the climb, level flight, descent, mixing state of different altitudes (two or more operation states of the same flight occur within the calculation time granularity), operation time, number of operations, and mixing coefficients of different altitudes are given. The indicator system is shown in Table 1.

Table 1 Indicator system

Dimension	Indicator	Abbreviation
Horizontal	Sector traffic flow	STF
	Variance of heading	VH
	Flight distance	FD
	Flight time	FT
	Average flight speed	AFS
Vertical	Time of climb	TC
	Time of descent	TD
	Time of level flight	TL
	Number of climb	NC
	Number of descent	ND
	Number of level flight	NL
	Time of climb in mixing state	TCM
	Time of descent in mixing state	TDM
	Time of level flight in mixing state	TLM
	Number of climb in mixing state	NCM
	Number of descent in mixing state	NDM
	Number of level flight in mixing state	NLM
	Mixing coefficient	MC

As for operation features, scholars have given calculation methods of some features from different aspects[18-19]. This paper uses ADS-B data. The density-based spatial clustering of applications with noise (DBSCAN) method is used to denoise the data. The indicator system provides methods of computation for features. Then we can calculate features based on it. Some definitions related to feature calculation are shown in Table 2, and some related algorithms are shown in Eqs.(1)—(3).

Table 2 Relevant definitions

Feature	Definition
Sector traffic flow	The number of aircraft included in a particular sector in the statistical period
Variance of heading	Variance of all flight directions entering the sector in the statistical period
Operating state	The status of flight climbing, level flight and descending
Mixing state	Two or more operating states occur in the same flight in the statistical period
Mixing coefficient	Mixing degree of aircraft operation state in statistical period
Climb state	The state that the aircraft climb rate is greater than the threshold value and the duration is greater than 30 s

Climbing time T_{lc} of the flight level l in the statistical period is

$$T_{lc}=\sum_{i\in l}t_{ilc} \tag{1}$$

where t_{ilc} is the climb time of flight i within 150 m above and below flight level l in the statistical period.

Number of climb N_{lc} of level l in the statistical period is

$$N_{lc}=\sum_{i\in l}n_{ilc} \tag{2}$$

where n_{ilc} is the number of climb of flight i within 150 m above and below flight level l in the statistical period.

In the statistical period, the mixing coefficient M_l of the operation of the flight level l is

$$M_l=\frac{N_{lc}\cdot N_{ll}+N_{ld}\cdot N_{ll}+N_{lc}\cdot N_{ld}}{N_{lc}+N_{ld}+N_{ll}} \tag{3}$$

where N_{lc}, N_{ld}, N_{ll} represent the number of climb, descent and level flight of flight level l in the statistical period, respectively.

1.2 Feature analysis and dimensionality reduction

The vertical range of area sector 5 in Central South of China is 4 500—30 000 m, which is the busiest area sector in Central South of China. In busy hours, it is usually divided into sector 5 (4 500—9 800 m) and sector 29 (9 801—30 000 m) with the same horizontal boundary, but the vertical boundary is not fixed. In order to identify the similar operation scenes of area sectors, this paper takes 4 500—13 100 m (actual flight level) of sector 5 as the research object and call it sector 5 below.

Based on the operation data of sector 5 in Central South of China from 1 October, 2018 to 31 December, 2018, feature values are calculated and analyzed using the indicator system in Table 1 .

Based on the indicator system in Table 1, 38 features are calculated, such as sector traffic flow, variance of heading, flight time, flight distance, average speed and number of climb, number of descent, etc. of FL4 500—FL13 100 m. Considering the large number of features and the possible correlation between data, information redundancy will be caused, which is not conducive to the identification of similar scenes. The correlation analysis of the calculated features is shown in Fig.3.

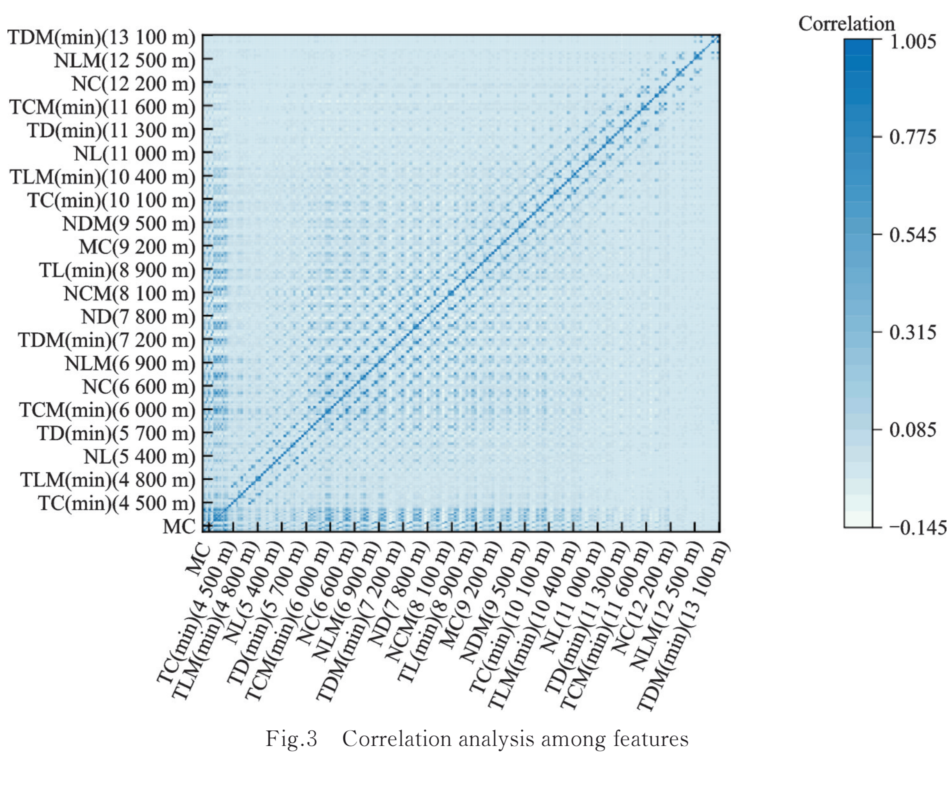

Fig.3 Correlation analysis among features

It can be seen that there is a high correlation between the horizontal operation features and the middle and low levels. Therefore, the operation of the sector in this area is of medium low levels clustering. At the same time, it can be seen that in Fig.3, the same features of each flight level on both sides of the diagonal are highly correlated. It shows that there is a large proportion of continuous climb and descent in the area sector.

Since there is correlation within features and the feature dimension is large, we need to use a mathematical method to reduce its dimension and correlation. PCA is a widely used method to reduce information redundancy and data dimension.

By means of an orthogonal transformation, PCA transforms the original random vector of correlation into a new random vector of uncorrelation.

Among them, the number of principal components (PCs) representing the minimum information loss is called effective dimension, and there are many principles widely used to determine effective dimension[20]:

(1) Parallel analysis finds the eigenvalues on a random dataset (noise of the same dimensions) and checks eigenvalues that are above noise levels.

(2) Kaiser rule drops all eigenvalues under 1.0 since these components contain more noise than signals.

(3) Acceleration factor checks the elbow in the scree plot, where there is a large drop in the eigenvalue between consecutive components. There is a significant gain in information by keeping the component just before the large drop. The number of components until the component just before the elbow is the recommended effective dimensionality from acceleration factor.

As shown in Fig. 4, with the increase of the number of the principal component, the cumulative variance of the principal component tends to be closer to 1, that is, the more information it contains. The change tends to be faster at the beginning and then slower. Using the principle of acceleration factor, when the cumulative variance is 0.85, the principal component contains more information, and the change slows down. It can be determined that the number of PCs is 60, and 60 of 382 features are used to measure similarity, so as to identify the operation scenes in area sector.

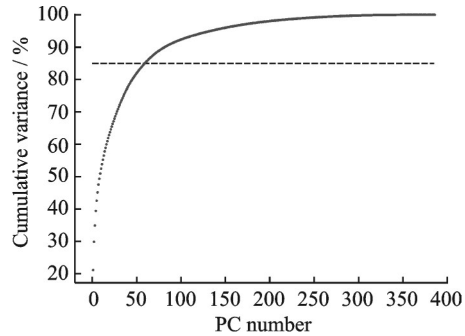

Fig.4 Cumulative variance explained vs number of PCs

Considering the numerous features and PCs, only three PCs are shown in Fig. 5. From distribution of their load coefficient, it can be seen that PC1 has a strong positive correlation with part of horizontal features (sector traffic flow, variance of heading), part of sector features (number of climb in the sector, time of level flight in sector), part middle and low level features (number of climb (4 500 m), number of level filght and descent (6 000—8 100 m)); and that PC2 has a strong positive correlation with part of horizontal features (flight time, flight distance), part low level descent features (number and time of descent (4 800—7 200 m)); while as a climbing PC, PC3 has a strong negative correlation with part of middle level descent features

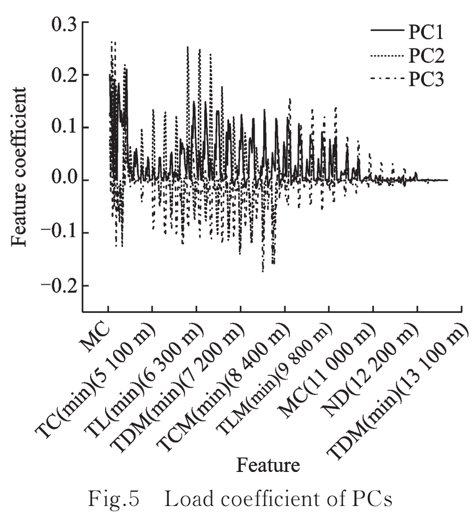

Fig.5 Load coefficient of PCs

(number and time of descent (5 700—8 100 m)) and a strong positive correlation with part of sector features (number of climb in secor, time of climb in sector), part high level climb features (number and time of climb (8 400—10 400 m)). Thus various PCs have climb, level flight, descent, horizontal, mixing attributes and this paper will analyse the results from various attributes.

2 Similar Scene Identification

In Section 1, PCA method is used to reduce the dimension of area sector operation features, and get 60 PCs. The samples with these 60 PCs need to be processed into discrete and sequential forms, and then their similarity needs to be measured to generate a measurement matrix. Next, the measurement matrix is put in the spectral clustering model to identify the similar scenes in the area sector.

2.1 Dynamic time warping

Dynamic time warping (DTW)[21] calculates the similarity between two time series by stretching or shrinking the time series. Compared with the European distance, it is looser for the sequence of time series. For two time series $X=\{x_1, x_2, \cdots, x_m\}$, $Y=\{y_1, y_2, \cdots, y_n\}$, the measurement principle is

$$\mathrm{DTW}(i,j)=\begin{cases} d(x_i,y_j)+\min\begin{cases}\mathrm{DTW}(i-1,j)\\ \mathrm{DTW}(i,j-1)\\ \mathrm{DTW}(i-1,j-1)\end{cases} & i\neq 0, j\neq 0\\ 0 & i=0, j=0\\ \infty & \text{Otherwise}\end{cases} \tag{4}$$

2.2 Spectral clustering

Spectral clustering algorithm[22-24] has become a new research hotspot in the field of machine learning in recent years. The spectral clustering algorithm is based on the spectral graph theory. Compared with the traditional clustering algorithm, it has the advantages of clustering on the sample space of any shape and converging to the global optimal solution. The principle is as follows:

(1) Based on kernel function, similarity matrix $\boldsymbol{W}$ is constructed by the ε-nearest-neighbor method or the k-nearest-neighbor method or the full connection method. v_i, v_j are two sample values, and under the Gaussian kernel function, the value w_{ij} in the similarity matrix is as follows

$$w_{ij}=\exp\left(-\frac{\left\|v_i-v_j\right\|_2^2}{2\sigma^2}\right)=-\exp\left(-\gamma\left\|v_i-v_j\right\|_2^2\right) \tag{5}$$

(2) Degree matrix $\boldsymbol{D}$

$$\boldsymbol{D}=\begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m\end{pmatrix} \tag{6}$$

$$d_i=\sum_{j=1}^{m} w_{ij}, w_{ii}=0 \tag{7}$$

(3) Laplace matrix $\boldsymbol{L}$ is constructed based on similarity matrix and degree matrix, and then $\boldsymbol{L}$ is standardized to get std.$\boldsymbol{L}$

$$\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{W} \tag{8}$$

$$\mathrm{std}.\boldsymbol{L}=\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{L}\boldsymbol{D}^{-\frac{1}{2}} \tag{9}$$

(4) The eigenvectors of the first k minimum eigenvalues are claculated, and input into a $m\times k$ matrix.

(5) Each row in matrix $\boldsymbol{F}$ is regarded as a k-dimension sample. There are m samples in total. Finally, a traditional method is selected for clustering.

2.3 Identifying similar scenes for area sector

Before measuring the similarity of data, considering the different information content of principal components, each principal component is weighed as

$$p_{\text{weight}}=\omega\cdot p_{\text{origin}} \tag{10}$$

where p_{weight} is the weighted PC, p_{origin} the original PC, and ω the proportion of information that can be explained by p_{origin}.

Controllers usually change shifts every 2 h, so the time granularity is set as 2 h. Using Euclidean distance and the DTW method, the time slice features of 15 min and eight consecutive time slices (2 h) are measured. Then based on Gaussian kernel function the similarity matrix $\boldsymbol{W}_{\mathrm{D}}$ of 8 832 × 8 832 discrete data and $\boldsymbol{W}_{\mathrm{T}}$ of 1 104 × 1 104 time series data are obtained.

Then, similar operation scenes for the area sector are identified with spectral clustering method.

(1) Since the Euclidean distance and the DTW method have been used to get the similarity measure matrices $\boldsymbol{W}_{\mathrm{D}}$ and $\boldsymbol{W}_{\mathrm{T}}$, they are directly used as input.

(2) The sum of each row in matrices $\boldsymbol{W}_{\mathrm{D}}$ and $\boldsymbol{W}_{\mathrm{T}}$ are calculated, that is, the degree matrix according to $d_i=\sum_{j=1}^{m} w_{ij}$ are calculated, and two diagonal matrices $\boldsymbol{D}_{\mathrm{D}}$ and $\boldsymbol{D}_{\mathrm{T}}$ are obtained.

(3) Using similarity measure matrices and degree matrices, Laplace matrices $\boldsymbol{L}_{\mathrm{D}}=\boldsymbol{D}_{\mathrm{D}}-\boldsymbol{W}_{\mathrm{D}}$ and $\boldsymbol{L}_{\mathrm{T}}=\boldsymbol{D}_{\mathrm{T}}-\boldsymbol{W}_{\mathrm{T}}$ are constructed. Then, the Laplace matrices are standardized to obtain std.$\boldsymbol{L}_{\mathrm{D}}=\boldsymbol{D}_{\mathrm{D}}^{-\frac{1}{2}}\boldsymbol{L}_{\mathrm{D}}\boldsymbol{D}_{\mathrm{D}}^{-\frac{1}{2}}$ and std.$\boldsymbol{L}_{\mathrm{T}}=\boldsymbol{D}_{\mathrm{T}}^{-\frac{1}{2}}\boldsymbol{L}_{\mathrm{T}}\boldsymbol{D}_{\mathrm{T}}^{-\frac{1}{2}}$.

(4) The eigenvectors of the first k_{D} and k_{T} eigenvalues of std.$\boldsymbol{L}_{\mathrm{D}}$ and std.$\boldsymbol{L}_{\mathrm{T}}$ are calculated respectively, and the matrices of $8\,832\times k_{\mathrm{D}}$ and $1\,104\times k_{\mathrm{T}}$ dimensions are formed.

(5) Matrix $\boldsymbol{F}_{\mathrm{D}}$ is considered as 8 832 samples with k_{D} features and matrix $\boldsymbol{F}_{\mathrm{T}}$ is considered as 1 104 samples with k_{T} features. Then K-means^{++} method is selected to cluster them to get clustering labels.

(6) According to the clustering label, the discrete data samples and the time series samples are divided into different classes to identify similar scenes of operation modes and operation trend.

3 Experimental Results and Analysis

In Section 2, the discrete and time series data are used to identify similar scenes of operation modes and trends in the area sector. However, the number of classes of similar scenes and the rationality of similar scene identification need to be determined by the analysis of results.

3.1 Clustering analysis

Average silhouette coefficient is a classical method to evaluate clustering results. It evaluates clustering results by calculating similarity between clusters and dissimilarity between different clusters. The formula for calculating the average silhouette coefficient s_i is[25]

$$s_i=\frac{b_i-a_i}{\max(a_i,b_i)} \tag{11}$$

where a_i represents the average distance between point i and all other points in the same cluster, and b_i the minimum of the average distance between point i and all other points in different clusters. Average silhouette coefficient is the average of all s_i.

3.2 Results of identification

3.2.1 Results of similar operation scenes

From Fig.6, it can be seen that the average silhouette coefficient of time series clustering results is 0.20 and 0.09 at the highest and the lowest, respectively while the average silhouette coefficient of discrete clustering is up to 0.41 when the number of clusters is 2. Based on the definition of the average silhouette coefficient, there is no obvious similar scenes of trend for the area sector with 15 min interval and 2 h length of single time series. However, the similar scenes of operation modes can be identified, and the best result is when the number of clus-

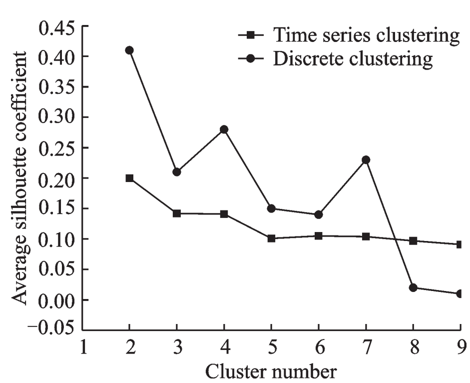

Fig.6 Average silhouette coefficient of clustering

ters is 2, that is, two types of similar scenes of operation modes can be identified. The ratio of cluster 1 to cluster 2 is 5 280∶3 552. Next, the original data will be matched with the labels of clustering result to analyze two different similar operation modes.

3.2.2 Comparative analysis of horizontal operation

Based on the indicator system in Table 1, four horizontal operating features, traffic flow, heading variance, average speed and flight time, are analyzed at first. Considering the measurement magnitude of different variables, all data are standardizedas 0—1 for visual analysis in Fig.7 and Table 3.

In Fig.7 and Table 3, it can be seen that the lower quartile of traffic flow value in operation mode 1 is larger than the upper quartile of traffic flow value in operation mode 2, and the mean of traffic flow in operation mode 1 is about 2.5 times of that in operation mode 2. However, the variance of traffic flow in mode 1 is lower than that in mode 2, and the traffic flow in mode 1 is more stable than that in mode 2. In terms of the total flight time in the sector, they are similar to the traffic flow. When traffic flow is low, pilots tend to fly at a more cost-effective speed, but because traffic is low, speed adjustments are relatively free. When traffic flow is high, controllers often use speed control strategies to regulate flights so that horizontal intervals between flights remain stable for easy control, but the speed can be adjusted in a small range. This also corresponds to the case where the average flight speed of mode 1 flight is lower than that of mode 2 and the variance of average flight speed of mode 1 is lower than that of mode 2 in Table 3. Meanwhile, mode 1 with higher flow rate has higher variance of heading and smaller variation due to similar flight states on different routes. Mode 2 with lower traffic flow has greater variation when the flight is on the same route and different routes.

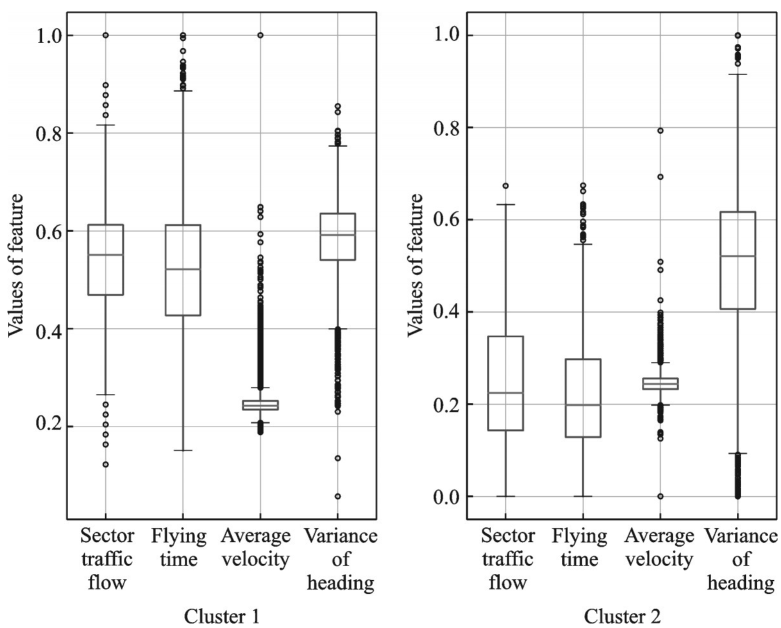

Fig.7 Analysis of horizontal operating features

Table 3 Variance of horizontal operating features in clustering

Variance	Sector traffic flow	Flight time/min	Average flight speed/(km•min^{-1})	Variance of heading
Variance of operating mode 1	0.011	0.018	0.001	0.006
Variance of operating mode 2	0.018	0.015	0.002	0.034

Horizontal operating feature analysis shows that operation mode 1 is a busy operation mode for the area sector and operation mode 2 is a relatively leisurely operation mode.

3.2.3 Comparative analysis of vertical operation

In order to further analyze the two operation modes of area sector, A comparison between the two operation modes is made from the perspective of vertical operation.

Level flight, descent and climb of each flight level in the two operation modes are analyzed, as shown in Fig.8. It can be seen that the curve of operation mode 1 are significantly higher than that of operation mode 2.

In terms of level flight operation, it can be seen from the number of operations that flights with level flight exist from low to high levels in both modes. Especially in mode 1, the number of level flight at the 6 000 m level is significantly higher than those of all other levels. However, the operating time indicates that flights of level flight mainly flying at high altitudes, explaining that due to speed control, duration of level flight at low and medium altitudes is short. At busy altitudes, the number and time of level flight of mode 1 are 3∶1 related to mode 2. At the same time, mode 1 fluctuates greatly at high altitudes, while mode 2 fluctuates slightly at different altitudes.

As far as descending operation is concerned, it mainly exists in medium and high levels. Curves of number and time are similar, indicating that the descent rate of flights is relatively stable. At the busy flight level, the number and time of descent of mode 1 are 4∶1 related to mode 2. Fluctuation between higher levels is greater in mode 1, while mode 2

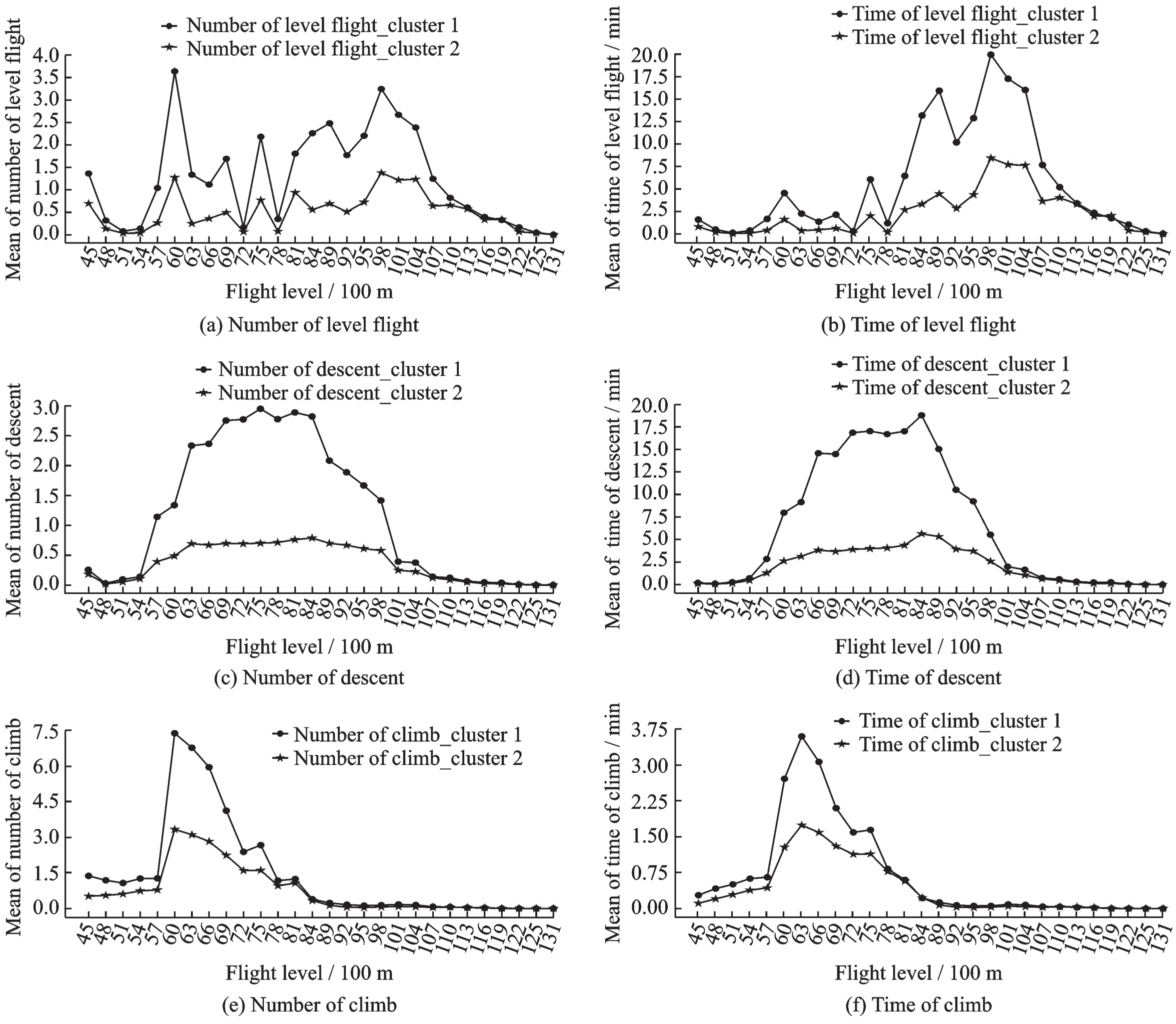

Fig.8 Operating number and time distribution at each flight level

shows a steady trend.

Climbing flights mainly exist at low level. The number and time of climb in mode 1 have a 2∶1 relationship with mode 2. The peak of operation number is at 6 000 m flight level, while the peak of operation time is at 6 300 m flight level. This shows that the 6 000 m flight level is the key level for the acceleration of climbing flights, and the initial climbing rate is low at 6 300 m.

In analysis of Fig.8, the 6 000 m level plays an important role in the climb, descent and level flight. As shown in Fig.9, the mixing coefficient of each level is further analyzed. Obviously, the distribution of mixing coefficient in the two operation modes is similar, and the peak is at the flight level of 6 000 m, which is greatly different from the mixing coefficient in the adjacent levels. The valley values are at 5 100, 7 200, 7 800 and 11 900 m, which can guide the dynamic management of division and combination for sector vertical operation.

From the above analysis, it can be concluded that operation mode 1 has great fluctuation at all levels, while operation mode 2 has stable operation at all levels. Compared with operation mode 2, operation mode 1 is more hybrid in all flight levels, and the busy degree of each flight level is higher than that of operation mode 2. In different operation modes, they have different approximate proportion relations. At the same time, the analysis results show that the 6 000 m level, especially in operation mode 1, plays an important role in the adjustment of climbing and descending flights. 6 000—7 500 m flight level plays an important role in the speed regulation for climbing flights. Therefore, in order to facilitate flight operation, it is necessary to avoid selecting 6 000 m and its adjacent levels as management boundary when the initiatives of sector formulating.

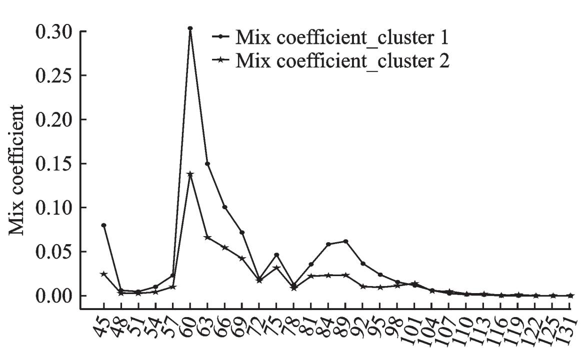

Fig.9 Mixing coefficient distribution at each flight level

4 Validation of Similar Operation Scenes

In order to verify the difference between the two operation modes, the time distribution of the two operation modes is analyzed. In Fig. 10, it can be seen that operation mode 1 is mainly distributed from 08:00 to 22:00, almost including all the time of early peak, noon peak and late peak; while operation mode 2 is mainly distributed between 23:00 to 07:00, which is also a relatively leisurely time. In peak hours the numbers of distribution of divison and mode 1 are similar, but before dawn they are

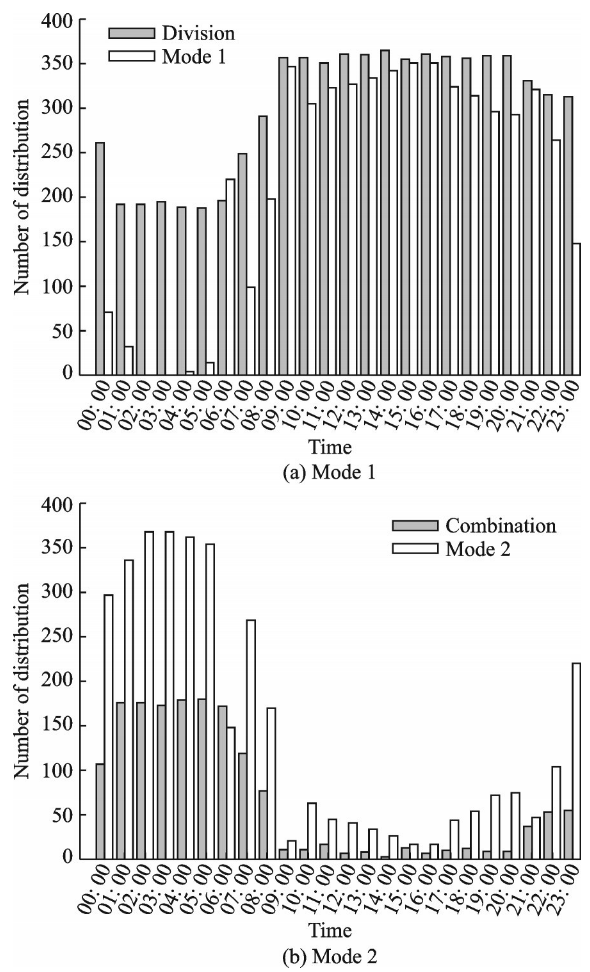

Fig.10 Time distribution of the two operation modes and actual operation

different. Moreover, although the trend of combination and mode 2 are similar, their amounts are different.

In order to find the reason for the above, the multidimensional scaling (MDS) method[26] is used in this paper, which maps the operating data of all time slices to two-dimensional space using the similarity measurement results between data. As shown in Fig. 11, it is the actual operation data, in which purple pentagons are the operation data of divison of sector and yellow triangles are the operation data of combination of sector. It can be seen that in the actual operation there are a lot of crossing parts. In other words, controllers make different decisions under similar conditions. It shows subjectivity of the actual operation.

Further, we analyze the duty roster of contollers and we find in leisurely time many vertical divison operations of the sector. It also lead to the amount of combination in leisurely time drops. It might be related to operation continuity, but we do not find its regularity.

Based on the above analysis, due to the subjectivity of actual operation, only the peak hours compliance of busy mode and division operation of sector are analyzed. It can be seen from Table 4 that at peak hours, the compliance rate of division operation of the sector and busy scene reaches 96.7%.

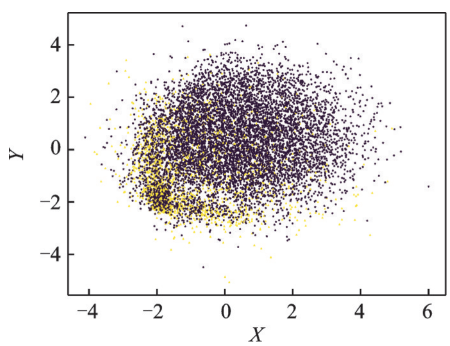

Fig.11 Distribution of actual operation

Table 4 Compliance of mode 1 and operation of division of sector in peak hours

Time	Busy scenes	Compliance	Compliance rate/%
Peak hour	4 426	4 280	96.7

Moreover, the distribution of cluster is shown in Fig.12, in which purple pentagons belong to mode 1 and yellow triangles belong to mode 2. It can be seen that there is no crossing part for the identification. It explains that for vertical operation, the result of identifacation is more objective than the actual operation. That shows the actual operation still needs improved. Meanwhile, it can also provide a objective initiatives for controllers.

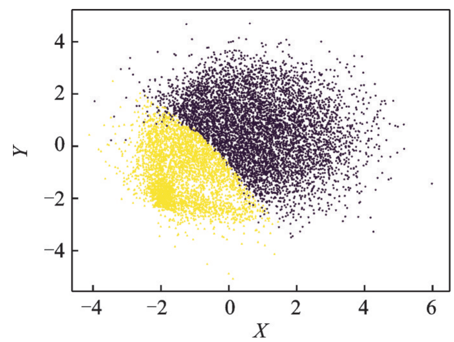

Fig.12 Distribution of cluster

From the above analysis, it can be concluded that for the area sector, the results of similar scenes of trend using time series features is poor. However, using the discrete-time slice features can effectively identify the similar scenes of the two operation modes. Operation mode 1 is a busy scene with a wider operation coverage, a higher degree of intersection and a greater volatility. Operation mode 2 is a relatively leisurely, stable operation scene. Finally, the actual operation of the area sector is used to verify it. The unreasonable operation at the leisurely time is found. Therefore, only peak hours validation is shown. At peak hours, the compliance rate of division operation of sector and busy scene reaches 96.7%. Also, It explains that for vertical operation, the result of identifacation is more objective than the actual operation. Thus the identification of similar scenes of operation mode is also of great significance to guide the actual vertical operation of the area sector.

5 Conclusions

Conserding the operation of the area sector, the operation indicator system reflecting the operation of area sector is established from the two dimensions of horizontal and vertical operation. Based on the indicator system, operation features are calculated. Due to the correlation between features, PCA method is used to reduce the dimension and information redundancy of data. Using acceleration factor principle, 60 PCs containing 85% information data are selected. Then Euclidean distance and DTW are used to measure the similarity of the selected PCs with discrete and time series forms. The measurement results are input into the spectral clustering model to identify the similar scenes of operation trend and the similar scenes of operation modes, and the clustering results are evaluated by the average silhouette coefficient. The clustering results show that under this complex feature and a 15 min time interval time sequence, there is no obvious similar scenes of trend for the area sector, but it can effectively identify similar scenes of operation modes. For the two similar scenes of operation modes identified, visual analysis and validation have been carried out. The analysis resluts prove the effectiveness of the identification of the operation modes. The analysis also show that actual operation is so subjective that we use the peak hours data to verify the results. The compliance rate of the busy scene identified in this paper is 96.7% with the actual division operation of the area sector at peak hours. Moreover, MDS figure shows there is no crossing part for the identification. It indicates that for vertical operation, actual operation still needs improve ment and the result of identifacation is more objective than the actual operation. It gives some suggestions on dynamic management of busy area sectors, and also provides a scene basis for the application of the other artificial intelligence techniques in air traffic control.

In this paper, we succussfully indentify the operation mode, but the result of indentification of operation trend is not obvious. Based on analysis in section 3, we will further modify the indicator system and try to explore operation trend scenes in different time intervals.

References

[1] SRIDHAR B. Applications of machine learning techniques to aviation operations: Promises and challenges[C]//Proceedings of 2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT). Tianjin, China: IEEE, 2020: 1-12.

[2] CHEN Y, BI H, ZHANG D, et al. Dynamic airspace sectorization via improved genetic algorithm[J]. Journal of Modern Transportation, 2013, 21(2): 117-124.

[3] SERHAN D, YOON S W, CHUNG S H. Dynamic reconfiguration of terminal airspace during convective weather: Robust optimization and conditional value-at-risk approaches[J]. Computers & Industrial Engineering, 2019, 132: 333-347.

[4] TANG J, ALAM S, LOKAN C, et al. A multi-objective evolutionary method for dynamic airspace re-sectorization using sectors clipping and similarities[C]//Proceedings of 2012 IEEE Congress on Evolutionary Computation. [S.l.]: IEEE, 2012: 1-8.

[5] YOUSEFI A. Optimum airspace design with air traffic controller workload-based partitioning[D]. Fairfax, USA: George Mason University, 2005.

[6] KUHN K D. A methodology for indetifying similar days in air traffic flow management initiative planning[J]. Transportation Research Part C, 2016, 69: 1-15.

[7] CHRISTIEN R, BENKOUAR A, CHABOUD T, et al. Air traffic complexity indicators & ATC sectors classification[C]//Proceedings of the 21st Digital Avionics Systems Conference. Irvine, USA: IEEE, 2002: 2D3.

[8] ANDREEVA-MORI A, MATAYOSHI N. Operational concept of traffic pattern classifier for optimal ground holding[C]//Proceedings of Air Traffic Management Research and Development (ATM Seminar). Vienna, Austria:[s.n.], 2019.

[9] BRINTON C R, PLEDGIE S. Airspace partitioning using flight clustering and computational geometry[C]//Proceedings of 2008 IEEE/AIAA 27th Digital Avionics Systems Conference. St. Paul, USA: IEEE, 2008.

[10] GORRIPATY S, LIU Y, HANSEN M, et al. Identifying similar days for air traffic management[J]. Journal of Air Transport Management, 2017, 65: 144-155.

[11] ARNESON H, BOMBELLI A, SEGARRA-TORNÉ A, et al. Analysis of convective-weather impact on pre-departure routing decisions for flights traveling between Fort Worth Center and New York Air Center[C]//Proceedings of the 17th AIAA Aviation Technology, Integration, and Operations Conference. Denver, USA: AIAA, 2017.

[12] GRABBE S R, SRIDHAR B, MUKHERJEE A. Similar days in the NAS: An airport perspective[C]// Proceedings of Aviation Technology, Integration & Operations Conference. Los Angeles, USA: AIAA, 2013: 4222.

[13] GRABBE S R, SRIDHAR B, MUKHERJEE A. Clustering days and hours with similar airport traffic and weather conditions[J]. Journal of Aerospace Information Systems, 2014, 11(11): 751-763.

[14] LIU Y, SEELHORSt M, POZDNUKHOV A, et al. Assessing terminal weather forecast similarity for strategic air traffic management[C]//Proceedings of the 6th International Conference on Research in Air Transportation. Istanbul, Turkey: [s.n.], 2014.

[15] VARGO E, TAYLOR C P. Similarity scoring with random field models for traffic flow management applications[C]//Proceedings of 2018 Aviation Technology, Integration, and Operations Conference. Atlanta, USA: AIAA, 2018: 3671.

[16] GORRIPATY S, HANSEN M, POZDNUKHOV A. Decision support framework to assist air traffic management[C]//Proceedings of 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC). Sacramento, USA: IEEE, 2016: 1-6.

[17] ASENCIO M. A clustering approach for analysis of convective weather impacting the NAS[C]//Proceedings of 2012 Integrated Communications, Navigation and Surveillance Conference. Herndon, USA: IEEE, 2012: N4-1-N4-11.

[18] LIU Y, HANSEN M. Evaluation of the performance of ground delay programs[J]. Journal of the Transportation Research Board, 2014, 24(2): 54-64.

[19] PRANDINI M, PUTTA V, HU J. A probabilistic measure of air traffic complexity in 3D airspace[J]. International Journal of Adaptive Control and Signal Processing, 2010, 24(10): 813-829.

[20] RAÎCHE G, WALLS T A, MAGIS D, et al. Non-graphical solutions for Cattell' s scree test[C]//Proceedings of International Meeting of the Psychometric Society. Montreal, Canada: IMPS, 2006: 1-12.

[21] HSU C J, HUANG K S, YANG C B, et al. Flexible dynamic time warping for time series classification[J]. Procedia Computer Science, 2015, 51: 2838-2842.

[22] NG A Y, JORDAN M I, WEISS Y. On spectral clustering: Analysis and an algorithm[C]//Proceedings of the 14th International Conference on Neural Information Processing Systems. Cambridge, USA: AIAA, 2002: 849-856.

[23] SINGH A. Spectral clustering: Machine Learning 10-701/15-781[R]. Pittsburgh, USA: Carnegie Mellon University, 2010.

[24] CHEN W, FENG G. Spectral clustering: A semi-supervised approach[J]. Neurocomputing, 2012, 77(1): 229-242.

[25] ROUSSEEUW P J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis[J]. Journal of Computational and Applied Mathematics, 1987, 20: 53-65.

[26] BORG I, GROENEN P J F. Modern multidimensional scaling: Theory and applications[M]. UK: Springer Science & Business Media, 2005.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Nos. 71731001, 61573181, 71971114) and the Fundamental Research Funds for the Central Universities (No. NS2020045).

Author Prof. **HU Minghua** is currently a professor and doctoral supervisor at College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China. He is the director of National Key Laboratory of Air Traffic Flow Management of China. His research interests include air traffic flow management and air space management.

Author contributions Prof. **HU Minghua** designed the study. Mr. **ZHANG Xuan** complied the models, conducted the analysis, interpreted the results and wrote the manuscript. Dr. **YUAN Ligang** collected data of the study. Dr. **CHEN Haiyan** contributed to the background of the study. Mr. **GE Jiaming** realized part of the algorithms. All authors commented on the manuscript draft and approved the submission.

Competing interests The authors declare no competing interests.

(Production Editor: ZHANG Bei)

面向繁忙区域扇区动态管理的相似运行场景识别

胡明华[1]，张 旋[1]，袁立罡[1]，陈海燕[2]，葛家明[2]

（1.南京航空航天大学民航学院/飞行学院，南京 211106，中国；

2.南京航空航天大学计算机科学与技术学院/人工智能学院，南京 211106，中国）

摘要：由于空中交通的不确定性，管制员在策略制定时面临着很大的挑战，而相似运行场景识别是一种很好的辅助管制员进行策略制定的方法。基于典型繁忙区域管制空域的运行特征建立了复杂度度量指标体系，在此基础上分析出区域扇区运行特征具有聚集性及连续性的特点，利用主成分分析有效地降低了数据维度和信息冗余，并利用主成分构建了代表运行模式场景和运行趋势场景的离散特征和时序特征。基于高斯核函数，采用欧氏距离和动态时间规整（Dynamic time warping, DTW）方法对特征间的相似度进行了度量，将度量结果输入到谱聚类模型中得到场景识别结果。聚类结果表明，基于上述指标体系，相似运行趋势场景识别效果不明显，相似模式场景识别结果较理想。最后采用多维缩放（Multidimensional scaling, MDS）方法对相似模式场景识别结果与扇区实际垂直运行进行了可视化对比，识别结果在高峰时刻能很好的反映运行情况，高峰时刻繁忙运行模式和开扇运行的匹配率达到96.7%，并分析出凌晨时段管制员在相似的场景下会做出不同决策，实验结果表明了识别结果的客观性及实际运行的主观性。相似的空中交通活动为管制策略制定提供了规律性支撑，也证明了这种方法在管制运行中对其他人工智能技术及动态策略制定的支持潜力。

关键词：空中交通；相似场景；无监督聚类；动态运行；时间序列；相似性度量