Dynamic Estimates of Sector Capacity Using the Combined SOM and K-means Method

Shuxuan Xiong[a], Yifei Zhao[b]*

Tianjin Key Laboratory for Air Traffic Operation Planning and Safety Technology, Civil Aviation University of China, Tianjin, 300300, P. R. China

[a]xiongshuxuan@hotmail.com, [b]yifei6666@sina.com

Abstract. It is well known that the sector capacity is particularly important for the air traffic flow management. In order to improve the accuracy in dynamic estimates of the sector capacity, a combined SOM and K-means method of clustering is proposed for the estimates in this paper. Firstly, the traffic flow data in term of the number of aircrafts within a time interval and a specific sector are acquired by using the outputs of the surveillance radar. The data are refreshed by sliding a window. The traffic flow data are then clustered by using the combined SOM and K-means method. The clustered results are compared with the results obtained by using the pure K-means method. The clustered results with high accuracy are suggested to be used. The sector capacity for each pattern is hence obtained by calculating the maximum flow for each sector. In the similar way, the late-coming data are matched with the patterns. The dynamic estimates of the sector capacity are hence completed. Using the real radar data we have performed the dynamic estimates of the sector capacity by following the above processes. The results show that the combined SOM and K-means method is efficient in the dynamic estimates of the sector capacity and may be useful for air traffic management.

Introduction

At present the air traffic flow increases dramatically as the civil aviation in China develops. However, the airspace capacity becomes to be a hurdle of the air traffic requests. The management of the air traffic flow is and important technology for solving this problem [1,2]. The efficient management of the air traffic flow is beneficial for the safe and efficient operation of air transportations. This paper focuses on the method of the sector capacity estimate, which is a key technology of the management of the air traffic flow.

So far several methods have been developed for the sector capacity estimates. These methods can be categorized into two classes. For the first class the sector capacity is estimated according to the controller workload [3-5]. These are subjective methods. The sector capacity is not estimated in real time. For the second class the sector capacity is estimated according to the characteristics of the traffic flow data acquired by the surveillance radar [6,7]. Using these methods the sector capacity can be estimated in a relatively objective way. Ref. [8] has discussed such methods in details. The number of aircrafts within 15min has been used for the traffic flow clustering. The sector capacity is hence estimated.

Following the basic ideas described in ref. [8], a combined SOM and K-means method is proposed for the sector capacity estimates. The contributions of the work are as follows. (1) The traffic flow data involved in the clustering are refreshed by sliding a window with flexible size. The real-time property of the method is hence guaranteed. (2) The traffic flow data are roughly clustered firstly by using the SOM method. The outputs of the SOM method are used as the initial value of the clustering using the K-means method. Using the combined method we have overcome the low accuracy property of the SOM clustering. And at the same time the initial value of the K-means clustering is determined. This is important because that a poor initial value will lead to poor accuracy of the K-means clustering. (3) We achieve the high accuracy of the clustering by using both pure K-means method and the combined method. The clustering patterns with high clustering accuracy are used as the final

clustering results. The real-time property and the high accuracy of the proposed method are shown by implementing the estimates of the sector capacity using the real radar data.

Dynamic Estimates of the Sector Capacity

Sector Capacity and Traffic Complexity

In air traffic flow management sector capacity is defined by the maximum number of aircrafts that can be operated by the controller provided that the controller's workload allows. Sector capacity is affected by the traffic complexity because that the traffic complexity will affect the controller workload. The controller workload will increase as the traffic complexity increases. The sector capacity will decrease correspondingly, and vice versa. The relationship of the sector capacity, traffic complexity, and the controller workload is demonstrated in Fig.1.
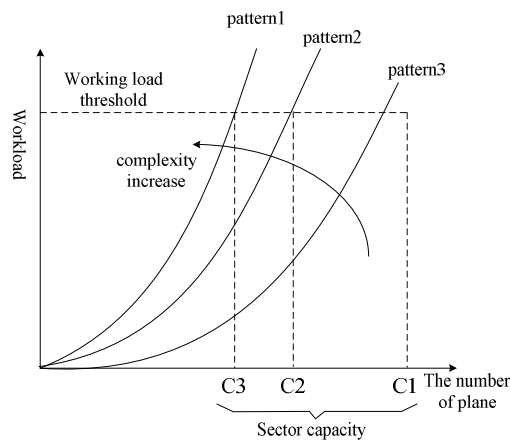

Fig. 1 relationship between traffic pattern, traffic complexity, controller workload and sector capacity

Estimating the Sector Capacity

The problem of the sector capacity estimates is that the sector capacity can not be expressed explicitly as a function of the controller workload and traffic complexity. In this paper the sector traffic density is used to describe the traffic complexity. The traffic flows are then clustered using statistics based on the traffic complexities. Several patterns of traffic flows are obtained as the clustering results. The sector capacity is obtained by calculating the maximum flow for each pattern. The refreshed data are clustered in the similar way. The process of estimating the sector capacity is detailed as follows, shown in Fig.2.
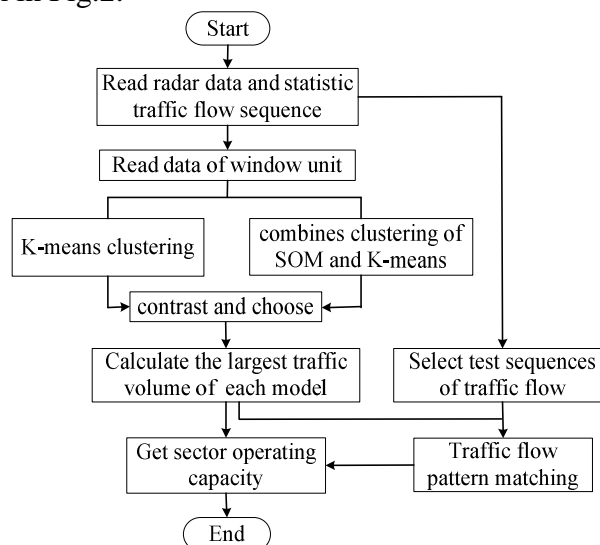

Fig. 2 the estimates of the sector capacity

(1) The traffic flow data are obtained by calculating the number of aircrafts for each sector within 15min according to the surveillance radar outputs.
(2) The traffic flow data are dynamically selected by sliding a window.
(3) The traffic flow data are clustered using the combined SOM and K-means method.
(4) The traffic flow data are also clustered by using the pure K-means method. These clustering results are compared with the results obtained using the combined method. The clustered patterns produced by the method with higher clustering accuracy are regarded as effective results. For each pattern the maximum flow is calculated for each sector.
(5) The sector capacity is obtained correspondingly.
(6) The refreshed traffic flow data are processed in the similar way. The dynamic estimates of the sector capacity are completed.

Definitions and Algorithms

Traffic Density and Traffic Flow

Sector traffic density is defined by the number of aircrafts for a specific flight segment of unit length. The traffic density is defined as

$$K = N / L , \tag{1}$$

where N is the total number of aircrafts for the flight segment, and L is the length of the flight segment.

The traffic flow, often in number of aircrafts per hour or per 15min, is defined by the number of aircrafts passing through the specific navigation or conjunction points of enroutes within the unit time. The traffic flow is expressed as

$$Q = N / T \tag{2}$$

Sliding Window for the Selection of Traffic Flow Data

The traffic flow data with different traffic complexities are obtained by calculating the number of aircrafts within 15min for the significant flight segments from the radar outputs, which are normally updated every 4sec. In this paper the traffic flow data in consideration are refreshed by using a sliding window. Data inside the window are involved in the clustering and data outside the window are not in our considerations. Furthermore, we assume data inside the window have the same contributions (weights) in the clustering. A window length of W denotes that the number of data involved in the clustering is W. W can be determined in simulations.

Selecting traffic flow data using a sliding window is shown in Fig.3. For example shown in the figure, as soon as four new data fall into the window, four old data are eliminated, and the clustering of the traffic flow has to be implemented. The dynamic estimates of the sector capacity are further realized.
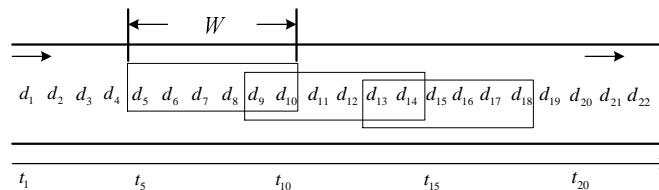

Fig. 3 a sliding window for the data selection

Combined SOM and K-means Method of Clustering

In this paper we propose the combined SOM and K-means method of clustering the traffic flow data. Using the proposed method the shortcomings of individual SOM and K-means method can be overcome. The characteristics of the three methods are detailed as follows.

K-means Method

K-means clustering algorithm developed three decades ago is one of the most popular clustering algorithm used in variety of domains. A priori knowledge of number of clusters are must for K-means clustering algorithm. K-means algorithm calculates its centers iteratively. A data set having k clusters, a set of k centers and the set of samples that belong to the kth cluster have to be created. K-Means algorithm minimizes the cost function denoting the distances between a pattern and and its cluster center.

K-means does not guarantee unique clustering because we get different results with randomly chosen initial clusters. Machine learning practitioners find it difficult to rely on the results thus obtained. The K-means algorithm gives better results only when the initial partitions are close to the final solution.

SOM Method

The Kohonen's Self-organizing Map (SOM) is a self-organizing neural network in which learning is unsupervised, which means that the only input to the network is the input pattern [9,10]. Moreover, learning is competitive, that is, neurons in the network compete for their activation induced by the input pattern and the winner adjusts its weights to incorporate information about the new input. The competition is performed by means of a similarity matching between the input pattern and the weight vectors. The topology of a SOM network is shown in Fig.4.

The SOM method is simple to implement. However, the clustering accuraccy of the SOM method is not satisfied and for relatively high clustering accuracy plenty of time is necessary for the convergence.
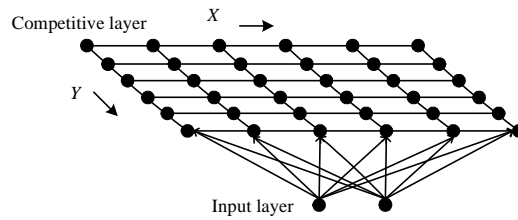

Fig.4   the topology of a SOM network

a Combined SOM and K-means Method for Clustering

In order to overcome the convergence problem of the SOM method and the inititialization problem of the K-means method. We have proposed the combined SOM and K-means method for clustering the traffic flow data. The concept is as follows.

Firstly the traffic flow data are roughly clustered using the SOM method. In this scenario the number of the learning times does not need to be large. One can set it to be 100, for example. At the end of learning the network nodes of the output layer become the neurons sensitive only to the certain pattern. The weight vector of each node becomes the center of the input pattern. Then the K-means clustering is further implemented, regarding the previous center vector as the initial cluster center. The clustering results are hence obtained using the combined SOM and K-means method. These results are to be evaluated afterwards.

In the K-means method the Euclidean distance is used to measure the difference between the samples, and the squared-error distortion for the difference between the patterns. The Euclidean distance of the sample x and y is defined by the square root of the sum of the squares of the p variables of the sample, expressed in the following equation.

$$EUCLID(x, y) = \sqrt{\sum_{i=1}^{p}(x_i - y_i)^2} \tag{3}$$

For the learning in SOM method, the learning coefficient $h(t)$ decreases from the initial value of $h(0)$ with the learning time.

$$h(t) = h(0)(1 - t/T) \qquad (4)$$

where $h(0)$ can be a relatively large value, t is the present iteration times, and T is the predetermined total iteration times.

In the similar way the size of the neighbour set descreases with the learning time as

$$S_i(t) = S_i(0)(1 - t/T) \qquad (5)$$

where $S_i(0)$ is the size of the initial neighbour set.

The weight denoting the connection between the output note i and its neighbour neuron is modified as follows
.
$$W_{ij}(t+1) = W_{ij}(t) + h(t) \left[x_j(t) - w_{ij}(t) \right] \qquad (6)$$

where $i \in S_i(t)$, $1 \le j \le n$, and $h(t)$ ($0 < h(t) < 1$) is the learning coefficient, which decreases to zero with time.

Selection of the Clustering Results

The clustering of the traffic flow data is also implemented by using the pure K-means. The corresponding results are compared with those obtained by using the conbined method. The cluster patterns obtained by the method with higher clustering accurary are used in the further work.

Results of the Dynamic Estimate of the Sector Capacity

In order to validate the proposed combined clusting method, we have estimated the sector capacity using the method and the acutual traffic flow data. Our data source is the outputs of the surveillance radar. The traffic flow data file is generated by the software of statisctics. The clustering using the combined method is implemented based on the R language.

Data Source

We firstly divide the flight path inside the sector with relatively large traffic flow into flight segments. The traffic density is eveluated for each flight segment. For the traffic flow data of an airport in the east of China, we divide the flight path of the sector into sevel segments. The traffic density is calculated for each 15min and each segment, denoted as E_1、E_2、E_3、E_4、E_5、E_6、E_7 in number of aircrafts per segement length. We use Q denoting the traffic flow in number of aircrafts per 15min. The data are based on flights for 8:00-20:00 within three days [11]. We have acquired 144 data samples. The nomalized traffic flow densities are shown in Table 1.

Table 1 data of traffic flow density

sample data	Traffic flow density							Traffic flow
	E_1	E_2	E_3	E_4	E_5	E_6	E_7	Q
1	-0.2879	0	0.9013	-0.3818	1.6276	1.3126	0.3368	-0.2502
2	-0.2879	-1.4396	-0.9013	-0.3818	1.6276	0.0135	0.3368	-0.9568
3	0.6324	-1.4396	-0.9013	-1.0868	0.8616	0.6631	0.3368	-0.6035
4	-1.2071	0.7198	-0.3004	-1.0868	0.6701	-0.6360	-0.3098	-0.6035

Clustering Performances

We have implemented two methods for clustering of the traffic flow density. The first method is the pure K-means method and the second is the combined SOM and the K-means method. The clustering performances of the two methods, indicated by the pattern square-error distortion divided by the total square-error distortion, are detailed in Table 2.

Table 2 the comparison of clustering performances

algorithm	The number of clustering types used				
	four	five	six	seven	eight
K-means	41.65%	46.21%	51.82%	57.41%	61.06%
combined method	45.42%	49.42%	54.89%	57.75%	61.88%

It is seen in the table that the clustering performance of the combined method is normally better than the pure K-means method. However, this situation may change when the source data are varied. Therefore, we suggest one may use the clustering results of the method which is shown to have better clustering performance.

Dynamic Estimates of the Sector Capacity

According to the clustering results, we calculate the maximum traffic flow for each pattern. The sector capacity is hence obtained as the maxiumum traffic flow for each sector. The cluster centers and the sector capacities are given in Table 3,4,5 for four, five, and six patterns. It is seen in the tables that for four patterns situation the clustering results are not so satisfied, while the clustering results for five and six patterns are more reasonable, which needs to be studied in the further work. In our application we use the five patterns for the consequent sector capacity estimates.

After determining the cluster centers of the five patterns, we obtain the sector capacity by the pattern matching. The pattern matching is performed by finding the pattern whose center is the nearest to the new coming data.

Table 3 results of normalized traffic flow density clustering (four patterns)

pattern	Traffic flow density							Traffic flow	Sector capacity
	E_1	E_2	E_3	E_4	E_5	E_6	E_7	Q	SC
1	-0.2874	0.1270	-0.4772	-0.4233	-0.4899	-1.0180	-0.9185	-0.4580	12
2	0.0409	-0.8740	-0.4292	-0.1300	0.4787	0.5238	0.6139	-0.5530	12
3	-0.4715	1.1516	0.6009	-0.099	-0.3638	0.4032	0.5954	0.6329	15
4	1.2898	0.2056	1.1589	1.4308	0.7522	0.8486	0.1520	1.314	18

Table.4 results of normalized traffic flow density clustering (five patterns)

pattern	Traffic flow density							Traffic flow	Sector capacity
	E_1	E_2	E_3	E_4	E_5	E_6	E_7	Q	SC
1	-0.0574	0.4198	-0.2503	-0.2056	-0.4787	-1.0691	-1.1182	-0.027	12
2	-0.6326	-1.1696	-0.9765	-0.3818	-0.1914	-0.5548	-0.0673	-1.1334	10
3	-0.7984	1.0397	0.3004	-0.2251	-0.6701	0.4465	0.4805	0.1422	13
4	0.6324	-0.5889	-0.3551	-0.0614	0.7921	1.017	0.5131	0.0066	14
5	0.7474	0.1799	1.5023	1.0281	0.5745	0.2571	0.4984	1.383	18

The dynamic estimates of the sector capacity are shown in Table 6, where the sample data used have been shown in Table 1. In Table 6 M_1、 M_2、 M_3、 M_4、 M_5 are the Euclidean distances of the four sample data and the cluster centers of the five patterns, respectively. PW denotes the determined pattern.

As soon as four new data enter into the sliding window, four old data are eliminated. The clustering has to be implemented and the sector capacity has to be estimated. In our application the size of the sliding window is 144.

Table 5 results of normalized traffic flow density clustering (six patterns)

| pattern | Traffic flow density | | | | | | | Traffic flow Q | Sector capacity SC |
	E_1	E_2	E_3	E_4	E_5	E_6	E_7		
1	-0.6818	-1.1311	-1.0730	-0.4825	-0.0136	-0.6360	-0.2174	-1.2091	8
2	-0.0574	0.4198	-0.2503	-0.2056	-0.4787	-1.0691	-1.1118	-1.0691	12
3	-0.9445	1.443	0.1287	-0.2811	-0.6701	0.3846	0.3368	0.2545	13
4	0.804	-0.8637	0.4206	0.0411	0.6701	0.4032	1.3715	-0.4622	10
5	0.6324	-0.4798	-0.5007	-0.0068	1.0318	1.1682	0.1931	0.1422	14
6	0.7474	0.1799	1.5023	1.0281	0.5745	0.2571	0.4984	1.3836	18

Table 6 results of dynamic sector capacity assessment

| sample data | Euclidean distance | | | | | PM | Sector capacity SC |
	M_1	M_2	M_3	M_4	M_5		
1	11.6961	11.8099	7.7573	3.6607	5.676756	4	14
2	9.5743	3.677783	13.3441	3.8295	12.62924	2	10
3	9.93086	4.768372	12.7749	2.203881	13.13392	4	14
4	2.4149	5.0899	2.5447	11.0231	14.1815	1	12

Conclusions

Based on the sector traffic density as an important indicator of the traffic complexity, we propose a combined SOM and K-means method for the dynamic estimates of the sector capacity. Using the combined method the high clustering performance is guaranteed. Additionally, the usage of the sliding window ensures us the real-time estimates of the sector capacity. The combined method has been used for dynamic estimates of the sector capacity using the real radar data. For the next step of the work we will investigate the estimates of sector capacity based on the other factor that may affect the sector capacity.

Acknowledgement

References

[1] Xiaohao Xu, Shan Li, Identification and prediction of air traffic congestion, Acta Aeronautica et Astronautic Sinica. 36 (2015) 2753-2763.

[2] Jerry D Welch, John Y N Cho, Ngaire K Underhill, et al. Sector workload model for benefits analysis and convective weather capacity prediction, 10th USA/Europe ATM R&D Seminar, 2013.

[3] J Y N Cho, J D Welch, N K Underhill, Analytical workload model for estimating en route sector capacity in convective weather, 9th USA/Europe ATM R&D Seminar, Berlin, Germany, 2011.

[4] A Klein, L Cook, Three models for weather impacted airspace capacity estimation and forecast, 9th USA/ Europe ATM R&D Seminar, Berlin, Germany, 2011.

[5] Jing Su, Area sector capacity evaluation based on multi factors influence, Nanjing University of Aeronautics and Astronautics, Nanjing, 2011.

[6] Lixia Song, Craig Wanke, Daniel Greenbaum, Predicting sector capacity for tfm, 7th ATM Seminar. Barcelona, Spain, 2007.

[7] Haijun Yan, Yong Tian, Lili Wang, A method of terminal airspace air traffic complexity, Aeronautical Computing Technique, 42 (2012) 63-67.

[8] Hongyong Wang, Wen Liu, Yifei Zhao, A calculation method of sector capacity in operation based on traffic flow pattern, Journal of Transportation Systems Engineering and Information Technology, 14 (2014) 188-193.

[9]T. Kohonen, Self-Organizing Map, Springer, New York, 1995.

[10] Tao Zhou, Huiling Lu, Clustering algorithm research advances on data mining, Computer Engineering and Applications, 48 (2012) 100-111.

[11] Hongyong Wang, Yifei Zhao, Fei Wang, Evaluation of air traffic control sector complexity, Journal of Transportation Systems Engineering and Information Technology, 13 (2013) 147-153.