Traffic & Transportation

Subjective Air Traffic Complexity Analysis Based on Weak Supervised Learning

Weining ZHANG[1], Weijun PAN[2], Changqi YANG[3], Xinping ZHU[4], Jianan YIN[5], Jinghan DU[6]

Original Scientific Paper
Submitted: 19 Apr 2024
Accepted: 17 Jan 2025

[1] wnzhang@cafuc.edu.cn, College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan, China
[2] wjpan@cafuc.edu.cn, Key Laboratory of Flight Techniques and Flight Safety, Civil Aviation Flight University of China, Guanghan, China
[3] 1132221610@qq.com, College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan, China
[4] zxp@cafuc.edu.cn, College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan, China
[5] j.yin@nuaa.edu.cn, College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[6] Corresponding author, dujinghan@cafuc.edu.cn, College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan, China

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT

Controller subjective evaluation is one of the most important ways to assess air traffic complexity. However, the inconsistency of human experts has a negative impact on the inference of complexity analysis models. To solve this problem, this paper proposes to construct a weakly supervised air traffic complexity dataset using highly reliable traffic situation similarity as labelling information. On this basis, a distance metric learning model is trained to generate a distance metric matrix that satisfies the similarity relationship. Finally, the K-means algorithm is combined to realise preferred complexity situation level classification and evolution analysis. Taking the actual operating data of a mid-southern area sector of China as an example, the effectiveness of the proposed method is verified. Experimental results show that the aircraft density, aircraft ground speed variance, heading disorder, convergence speed and horizontal conflict have a greater impact on the complexity situation. Compared with the K-means algorithm based on Euclidean distance, metric learning improves the optimal silhouette coefficient and Davidson-Boldin index by 31.80% and 12.97%, respectively. In addition, it is confirmed that the situation evolution is driven by one or two key influencing factors.

KEYWORDS

air traffic complexity; subjective similarity; weakly supervised; distance metric learning; cluster analysis.

1. INTRODUCTION

The continuous growth of air traffic demand drives the vigorous development of air transport industry. At the same time, it is accompanied by the emergence of airspace congestion, flight conflicts and other problems. In order to maintain proper separation between aircraft and ensure their safe, efficient and orderly operation, air traffic controllers need to monitor traffic situation in real time and issue control instructions to pilots [1]. However, when the traffic density reaches airspace capacity limits, controllers are in a high workload state, which may lead to operational errors and unsafe incidents. Air traffic complexity [2], as an indicator describing the difficulty of controllers in monitoring and managing traffic in the airspace they are responsible for, can reflect the control pressure faced by controllers to a certain extent. Accurately analysing air traffic complexity situation in advance, on the one hand, helps adjust traffic flow to achieve a balance between airspace capacity and demand; on the other hand, it also provides a reference basis for re-sectorising control sectors to realise dynamic airspace configuration.

How to analyse the complexity of air traffic has always been a hot issue for researchers. In earlier studies, the number of aircraft was used to describe the traffic situation in a given airspace. For example, the Enhanced Traffic Management System (ETMS) [3] achieves the purpose of monitoring and alerting airspace sectors by comparing the predicted future traffic flow with the predefined flow threshold. Recent studies have shown that air traffic complexity is often affected by multiple factors such as sector geometry, traffic flow, aircraft performance, potential conflicts, weather conditions etc [4]. Laudeman et al. [5] proposed the concept of dynamic density (DD) by linearly combining various potential influencing factors. Compared with traffic density, it has a higher correlation with air traffic complexity and is more interpretable. In addition, in order to capture the complex nonlinear relationships existing among influencing factors, nonlinear models such as neural networks are also widely used [6-7]. Essentially, these models use various influencing factors calculated from trajectory data as input to predict quantitative indicators of air traffic complexity, such as controller physical activity, subjective evaluation, sector status etc. In particular, the subjective evaluation of controllers has become one of the most important ways to measure air traffic complexity [8-10].

As with general supervised learning tasks, the prediction accuracy of air traffic complexity models relies on sufficient and reliably labelled datasets. However, obtaining such datasets often requires the active participation of domain experts and expensive time costs, such as air traffic controllers providing real-time feedback on subjective feelings about airspace operating conditions and workloads. Because of the above reasons, the collected datasets have limitations such as small sample size and low variability. In addition, due to the inconsistency of controllers' subjective evaluations, it is often difficult to ensure the accuracy of labelling information, thus causing interference to the learning process of the model [11]. To alleviate this problem, Cao et al. [12] introduced transfer learning concept into airspace operation complexity evaluation for small-training-sample environment. It fully mines and utilises the knowledge of other sectors to improve the prediction accuracy of target sector complexity. Li et al. [13] proposed deep unsupervised learning approach for airspace complexity evaluation, which directly circumvents the use of labelled samples. They introduced a new loss function to enhance the generalisation ability of the unsupervised model, and verified that the proposed model can obtain the best evaluation performance in six airspace sectors. In addition, Antulov-Fantulin [14] established a mathematical model to determine air traffic complexity based on air traffic controller tasks for the given traffic situations, which has been verified in en-route airspace. Similar, Jurinić et al. [15] proposed a new set of terminal airspace complexity indicators based on air traffic controller tasks and input. Pérez Moreno et al. [16] attempted to define which variables determine airspace complexity based on machine learning models. They found that the aircraft number, traffic flows occupancy and aircraft vertical distribution are the main indicators. On this basis, they further developed a dynamic complexity indicator and used machine learning models to predict its complexity [17].

Through extensive communication and discussion with frontline controllers, we found an interesting phenomenon, where most controllers believe it is easier to subjectively calibrate the complexity similarity (i.e. coarse-grained labels) between different airspace operation scenarios than to calibrate the complexity level (i.e. fine-grained labels) of each scenario. Based on the above phenomenon, we attempted to model the air traffic complexity analysis problem as a weakly supervised learning paradigm. Weakly supervised learning, as an important paradigm of machine learning, can make full use of inaccurate supervision information (i.e. similarity) to overcome the problems of high labelling difficulty and high noise from accurate categories and improve the generalisation ability of the model [18]. Inspired by the successful application of weakly supervised learning paradigm in various tasks, such as image retrieval, face recognition etc., this paper constructs an air traffic complexity dataset from a weakly supervised perspective and on this basis proposes a framework for the situation analysis of air traffic complexity. The framework fully combines distance metric learning and cluster analysis technology to provide rich insights in quantifying the importance of complexity factors, classifying complexity situation levels and revealing the causes of situation evolution. A case study of the proposed framework is conducted using real operation data from a mid-southern area sector of China. In general, the main contributions of this paper are as follows:

1) To the best of our knowledge, this paper is the first to introduce the weakly supervised learning paradigm to the problem of air traffic complexity analysis.
2) It provides a new way to identify the importance of complexity factors. By introducing distance metric learning methods, the importance of different complexity factors themselves and their interactions is quantified, and the key influencing factors that drive the change of situation level are given.
3) It presents a new method for determining the complexity level of air traffic. The proposed framework combines metric learning and cluster analysis to obtain the number of complexity levels that meet the controller's cognitive preferences, which to some extent alleviates the inconsistency (i.e. inconsistency within and between controllers) in the process of calibrating the accurate complexity level in past studies.

2. METHODOLOGY

The proposed analysis framework for air traffic complexity situation is shown in *Figure 1.* It first generates a weakly supervised dataset with airspace operation scene pairs as basic data items, where each scene pair includes its own complexity factors and similarity labels between the two. Potential influencing factors are calculated based on flight trajectory data and specific factor definitions, while similarity labels are derived from controllers' subjective evaluation of paired radar images. Based on the generated dataset, the distance metric learning model is trained to generate distance metric matrices to reflect the importance of the influencing factors themselves and the interactions between factors. Finally, using the distance metric matrix as prior information, a biased cluster analysis based on K-means algorithm is performed to realise the perception of air traffic complexity from two aspects: the basic levels of situation classification and the critical factors of situation evolution.

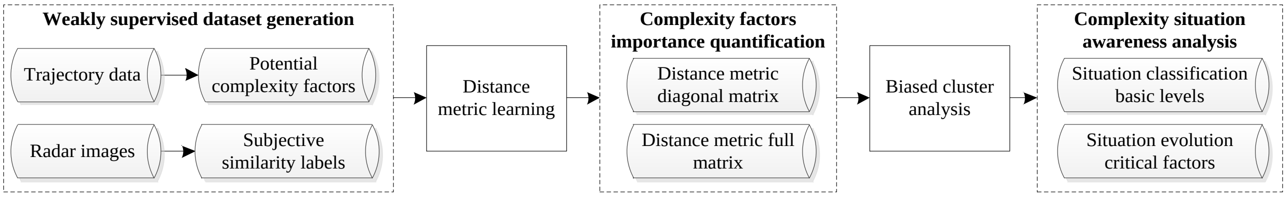


Figure 1 – The main framework for analysing air traffic complexity situation

2.1 Weakly supervised dataset generation

In order to obtain a weakly supervised dataset for the airspace sector evaluation, complexity factors are first constructed from three perspectives: traffic flow, aircraft performance and potential conflicts, as shown in *Table 1*. These complexity factors are derived from literature [6]. Due to its comprehensiveness and convenience, most mainstream studies on airspace complexity analysis and prediction continue to use this set of factors. For the consistency of comparative analysis, this paper adopts the same settings. Among them, traffic flow factors directly reflect the distribution of aircraft in the sector currently and in the future, and are usually used as basic factors to describe the airspace traffic situation in the practical application of air traffic control systems. Aircraft performance factors mainly include the speed parameters of aircraft operation while potential conflict factors quantify the risk of collision between aircraft. For example, the separation and convergence sensitivity factors describe the effects of changes in aircraft speed and heading on relative distances, and are important factors that cause a surge in controller workload. The calculation method of specific factors can be found in the literature [6] and will not be described in detail here. Specifically, we collected the real flight trajectory data of a mid-southern area sector of China from 12:00–13:00 on 1 to 7 December 2019. The selected data not only ensures the diversity of airspace operation scenarios, but also takes into account the time cost of ATCOs to label similarity information. Each data piece contains flight number, timestamp, position (longitude, latitude and altitude), speed and other information. On this basis, the above complexity factors were calculated with 1 minute as the basic time granularity. Further, the air traffic control radar images of the corresponding period (one image per minute, a total of 420 images) were extracted and submitted to the control experts for calibration of complexity similarity information. Just like the subjective complexity level calibration in mainstream literature, similarity information used in this paper is also based on the controller's subjective perception of the airspace situation.

Table 1 – Characterisation of air traffic complexity factors

Types (ID)	Variables	Meanings
Traffic flow (1-6)	N	Number of aircraft in the sector
	F_5, F_{15}, F_{30}, F_{60}	Number of aircraft entering the sector in the next 5, 15, 30, and 60 minutes
	S_{dens}	Aircraft density within sector
Aircraft performance (7-9)	σ_{gs}^2	Variance of aircraft ground speed
	σ_{gs}/v_{gs}	The ratio of the standard deviation of the aircraft ground speed to the mean
	v_{vs}	Aircraft mean vertical speed
Potential conflict (10-23)	S_{vpro_1}, S_{vpro_2}	Average and minimum vertical proximity
	S_{hpro_1}, S_{hpro_2}	Average and minimum horizontal proximity
	S_{track_dis}, S_{speed_dis}	Heading and speed disorder
	S_{div}, S_{conv}	Separation and convergence speed
	S_{sensi_d}, S_{insen_d}	Separation sensitivity and insensitivity
	S_{sensi_c}, S_{insen_c}	Convergence sensitivity and insensitivity
	S_{inter_vert}, S_{inter_hori}	Vertical and horizontal conflicts

The specific calibration process is shown in *Figure 2*. Specifically, two images are randomly selected from all air traffic control radar images by sampling with replacement and submitted to the controllers as a scene pair. Then, based on the experts' cognitive experience of the airspace traffic situation, we evaluate whether the two images are close in complexity. If there is no obvious difference, the scene pair is marked as similar; otherwise, it is marked as dissimilar. By repeating the above process multiple times, similar sets and dissimilar sets of complexity samples can be obtained. Considering the relevant regulations of the area control centre, four controllers are randomly selected to participate in the calibration process from the airspace to be evaluated. These four controllers are all third-level controllers (which represent highly competent and experienced controllers certified by the Air Traffic Management Bureau of China) responsible for the corresponding area sector. They all have more than 8 years of work experience and are between 40 and 50 years old. In order to ensure the reliability of the controller's labelling information, each scene pair is evaluated three times by the four experienced controllers, and scene pairs with consistent evaluation results are selected. In total, we calibrated 2,500 scene pairs, including 2,397 pairs with consistent evaluations (including 1,352 similar pairs and 1,045 dissimilar pairs) and 103 pairs with inconsistent evaluations. For those with inconsistent evaluation results, considering that they rarely occur, we exclude them from the dataset.

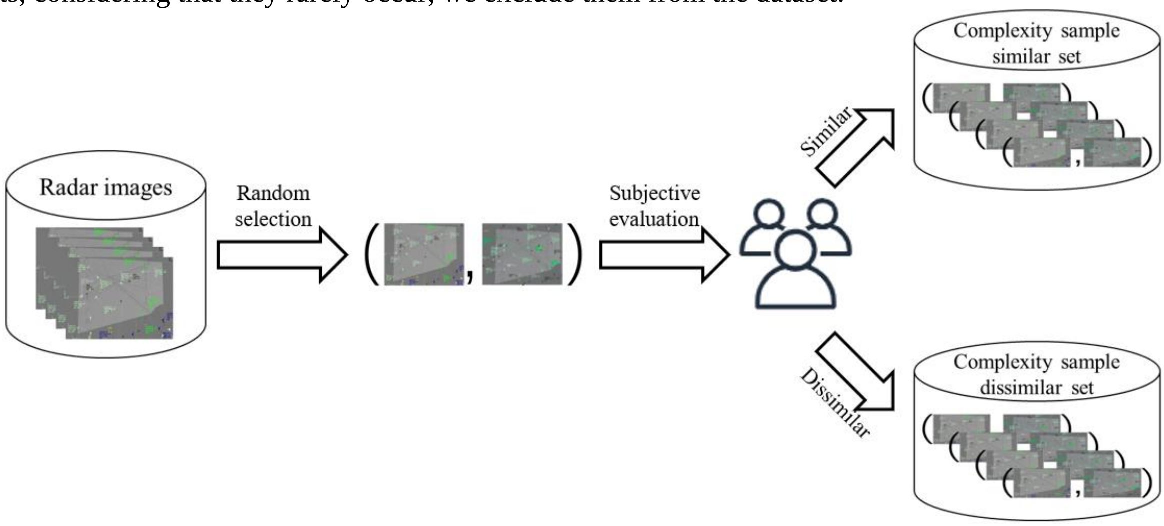


Figure 2 – Weakly supervised information labelling process for air traffic complexity samples

2.2 Complexity factors importance quantification

In order to make the distance metric between complexity samples comply with the similarity and dissimilarity relationships calibrated by the controllers, the importance of each factor needs to be accurately assessed. Traditional distance metric methods, such as Euclidean distance, essentially assume that different factors of complexity samples have the same importance, which may cause problems such as similar scene pair being too far away or dissimilar scene pair being too close. Distance metric learning (DML) [19] is one of the effective ways to alleviate this problem. It usually uses a given weakly supervised sample set to learn a distance metric matrix, so that in the new metric space, similar scene pairs are closer and dissimilar scene pairs are further apart. The learned metric matrix can also be applied to various general tasks, such as classification, clustering, feature extraction, information retrieval and others.

Specifically, this paper selects a classic distance metric learning algorithm based on pairwise constraints, namely the Mahalanobis metric for clustering (MMC) [20], to generate the distance metric matrix. MMC takes minimising the sum of squared Mahalanobis distance between pairs of similar scenes as the optimisation goal, and takes the sum of Mahalanobis distance between pairs of dissimilar scenes to be greater than a predefined threshold (usually 1) as a constraint to build an optimisation model. The specific model is as follows:

$$\min_{A \geq 0} \sum_{(x_i,x_j)\in SS} d_A^2(x_i, x_j) \qquad s.t. \sum_{(x_i,x_j)\in DS} d_A(x_i, x_j) \geq 1 \tag{1}$$

where $x_i \in R^{23}$ represents a sample composed of 23 complexity factors, and SS and DS are similar and dissimilar sets of complexity samples, respectively. $d_A(x_i, x_j) = \sqrt{(x_i - x_j)^T A (x_i - x_j)}$ is the Mahalanobis distance between x_i and x_j, and $A \in R^{23\times 23}$ is the distance metric matrix to be learned. $A \geq 0$ means that A is a positive semi-definite matrix to ensure that it satisfies the non-negativity and triangle inequality characteristics of distance metric. When A is the identity matrix, $d_A(x_i, x_j)$ degenerates into the Euclidean distance; when A is a diagonal matrix, it indicates that only the importance of the factors themselves is considered; when A is a full matrix, it indicates that the importance of both the factors themselves and their interactions is considered. Since this problem is a convex optimisation problem, it can be solved using the classic gradient descent algorithm.

2.3 Complexity situation awareness analysis

In order to further mine the cluster information contained in the complexity samples and reasonably divide the complexity levels, an improved K-means clustering algorithm based on the metric matrix is proposed to realise the classification of air traffic complexity levels under weak supervision information. Specifically, as a common cluster analysis technique, K-means clustering aims to minimise the sum of squared Euclidean distances between each sample and the cluster centre to which it belongs, so as to generate a partitioning scheme containing K clusters. Its optimisation goal is as follows:

$$\mathrm{f} = \sum_{i=1}^{N} \sum_{j=1}^{K} r_{ij} ||x_i - \mu_j||^2 \tag{2}$$

where N is the number of samples, K is the number of clusters, μ_j is the centre of cluster j. When sample x_i is in cluster j, r_{ij} is 1; otherwise it is 0. In contrast, the improved K-means clustering algorithm uses the well-learned Mahalanobis distance to replace the traditional Euclidean distance to find clusters that are more in line with the expected semantics (i.e. the similarity relationship of complexity samples). The specific optimisation goal is as follows:

$$\mathrm{f} = \sum_{i=1}^{N} \sum_{j=1}^{K} r_{ij} d_A^2(x_i, \mu_j) \tag{3}$$

where $d_A(x_i, \mu_j)$ is the Mahalanobis distance between sample x_i and the centre μ_j of cluster j under the metric matrix A. *Table 2* gives the detailed steps of the algorithm.

Table 2 – K-means clustering algorithm based on Mahalanobis distance

Input: complexity dataset $D = \{x_1, x_2, \dots, x_N\}$, the number of clusters K, distance metric matrix A
Output: complexity sample classification results $C = \{c_1, c_2, \dots, c_K\}$
1. Randomly select K samples from the complexity dataset as the center vector of the initial cluster $\{\mu_1, \mu_2, \dots, \mu_K\}$
2. Repeat:
3. Initialise $c_i = \{\}, i = 1,2, \dots, K$
4. For $i = 1,2, \dots, N$
5. Calculate the Mahalanobis distance $d_{ij} = d_A(x_i, \mu_j), j = 1,2, \dots, K$ between sample x_i and cluster μ_j
6. Determine the cluster λ_i to which sample x_i belongs based on $\lambda_i = argmin_{j\in\{1,2,\dots,K\}} d_{ij}$
7. Classify sample x_i into the corresponding cluster $c_{\lambda_i} = c_{\lambda_i} \cup \{x_i\}$
8. Update the centre vector of each cluster $\mu_i = \frac{\sum_{x\in c_i} x}{\lvert c_i \rvert}, i = 1,2, \dots, K$
9. Until the centre vectors of all clusters converge

With the determined traffic situation complexity level, a natural idea is to identify the important factors that cause the situation to evolve. Hereby, the normalised factor gap G_{ij}^k is defined to describe the degree of change in the k-th factor between normalised sample i and j. It is formalised as follows:

$$G_{ij}^k = (x_i^k - x_j^k) A_{diag}(k,k) (x_i^k - x_j^k) \tag{4}$$

where $A_{diag}(k,k)$ is the k-th element on the main diagonal of the learned distance metric diagonal matrix A_{diag}, x_i^k is the k-th factor of the normalised sample i. Using G_{ij}^k, we can directly compare the contribution of different factors to the distance between two situation samples.

3. CASE STUDY

3.1 Experimental settings

In the experiment, a total of 2,000 scene pairs are randomly selected from the generated dataset, including 1,000 similar scene pairs and 1,000 dissimilar scene pairs. The proposed algorithm is implemented using the distance metric learning framework metric_learn 0.6.2 [21] and the machine learning library Scikit-learn 0.22.2 [22]. In the parameter settings of the metric learning model, the maximum number of iterations is 100, the convergence threshold is 0.001, and the initialisation method is the unit matrix. On this basis, two types of metric matrices are generated, namely diagonal matrix and full matrix. As for the parameter settings of the K-means clustering algorithm, the number of clusters is set in the range of 2 to 20. To evaluate the performance of different clustering algorithms and determine the optimal number of clusters, the silhouette coefficient (SI) and the Davidson-Boldin index (DBI) are selected, which can be calculated as follows:

$$\mathrm{SI} = \frac{1}{N} \sum_{i=1}^{N} \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{5}$$

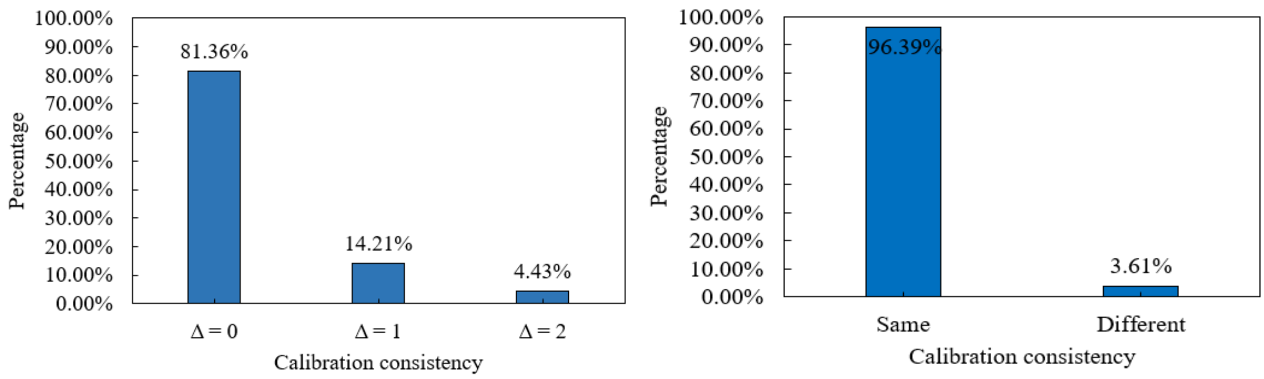
$$\mathrm{DBI} = \frac{1}{K} \sum_{i=1}^{K} \max_{j\neq i} \frac{c(i) + c(j)}{\lvert\lvert \mu_i - \mu_j \rvert\rvert_2} \tag{6}$$

where $b(i)$ is the minimum average distance between sample i and other clusters, $a(i)$ is the average distance between sample i and all other samples with the same cluster, $c(i)$ is the average distance of all samples in cluster i. Among them, the larger the SI, the better the cluster performance, and the opposite is true for DBI.

3.2 Analysis of weakly supervised sample effectiveness

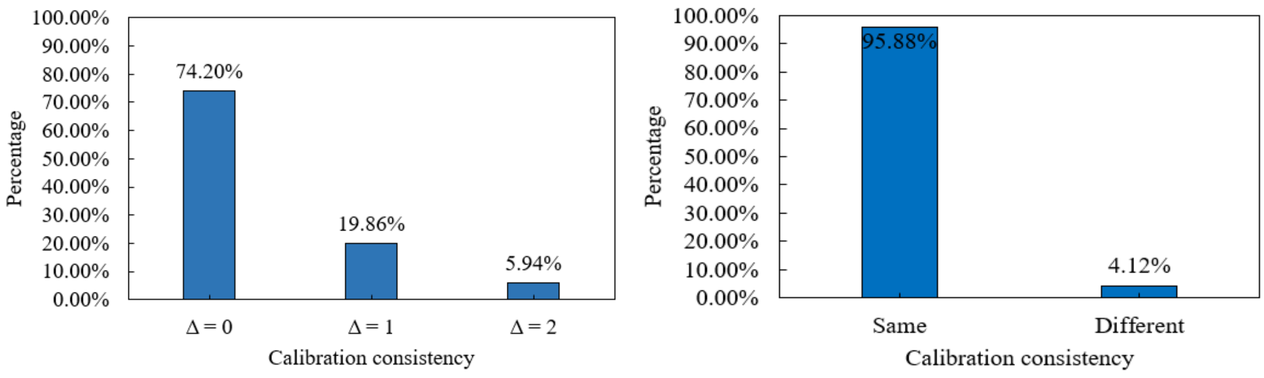
To compare the differences in calibration consistency between complexity level and complexity similarity, we further subjectively calibrated all radar images on a complexity level from 1 to 4, where each image was also evaluated three times by four controllers. A total of 5,040 calibration results were recorded (12 calibrations per image, a total of 420 images). *Figure 3* compares the consistency of strong supervision information (i.e. complexity level) and weak supervision information (i.e. complexity similarity) during multiple calibrations by a controller. It can be seen that the calibration consistency of strong supervision information is poor. There are not only many cases where the complexity level differs by 1, but also a small number of cases where the complexity level differs by 2. It can be expected that the difficulty of calibration will increase significantly when there are more complexity levels. In comparison, the calibration consistency of weakly supervised information is better, with only a few cases with different similarities.

Furthermore, *Figure 4* shows the subjective calibration consistency of different controllers. Through comparison, it can be found that the consistency of complexity levels decreases significantly, while the consistency of complexity similarity is less affected. This phenomenon directly reflects the advantages of weakly supervised information, which not only avoids the cognitive differences of different experts on different complexity levels, but also improves the availability of supervised samples to a certain extent, and obtains more reliable calibration results with less time cost.



(a) *(b)*

Figure 3 – Consistency analysis of subjective calibrations by the same controller: a) complexity level; b) complexity similarity



(a) *(b)*

Figure 4 – Consistency analysis of subjective calibrations by different controllers: a) Complexity level; b) Complexity similarity

3.3 Analysis of distance metric learning effectiveness

In order to verify the effectiveness of metric learning, the Mahalanobis distances of all similar scene pairs and dissimilar scene pairs were calculated respectively, and their distribution is shown in *Figure 5*. It can be seen that similar scene pairs are smaller than dissimilar scene pairs in terms of the median Mahalanobis distance. Furthermore, the two boxplots are completely separated in the region consisting of the lower and upper quartiles, with no overlap. This shows that the Mahalanobis distance of similar scene pairs is generally smaller than the distance of dissimilar scene pairs, which fully confirms the effectiveness of the metric learning model.

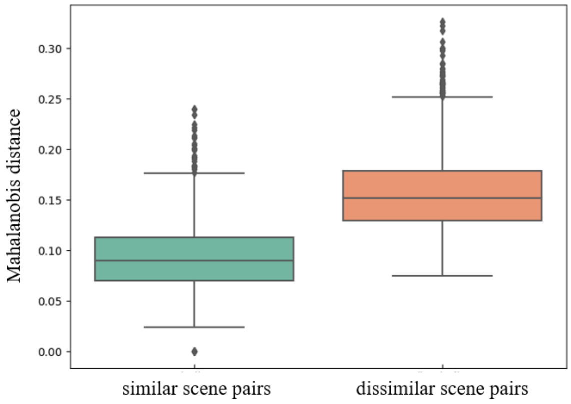
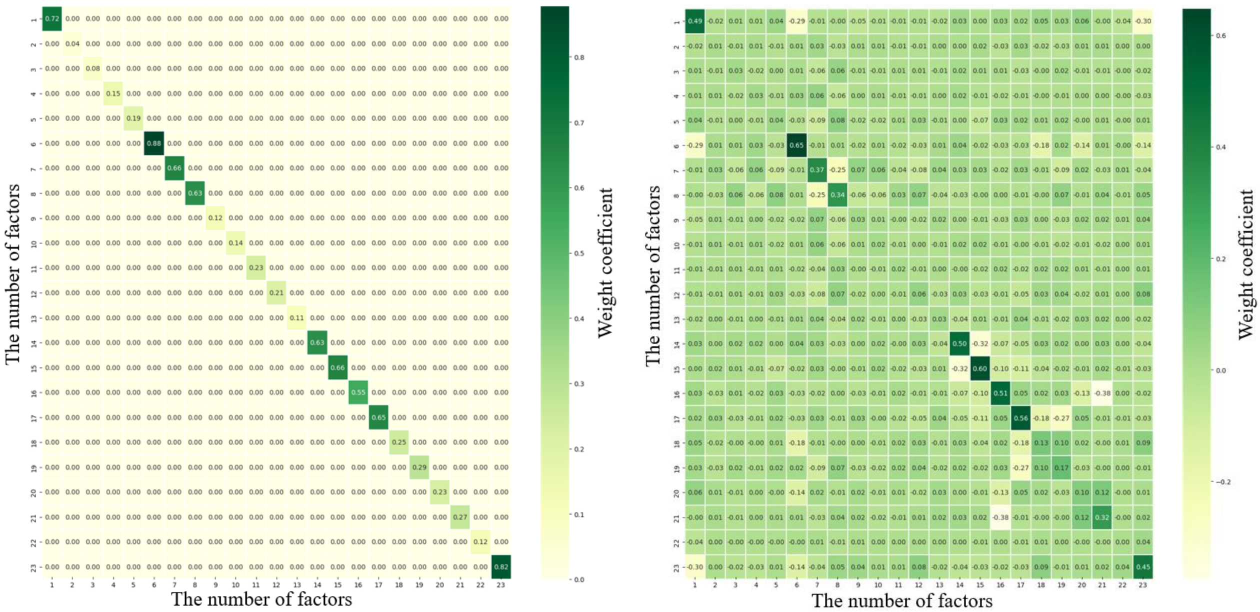


Figure 5 – Mahalanobis distance between similar scene pairs and dissimilar scene pairs

Figure 6 further visualises the two types of learned metric matrices, and the depth of their colours reflects the size of the weight coefficients. The elements on the diagonal are the weight coefficients of each complexity factor, and the elements on the off-diagonal are the weight coefficients of the interaction between the two factors. Since the complexity factor is normalised before metric learning, the size of the weight coefficient can reflect the importance of different factors and interactions between factors. *Figure 6(a)* shows the distance metric matrix in the form of a diagonal matrix, from which the key factors that affect the distance between two scenes can be summarised. Specifically, for traffic flow factors (1~6), factor 1 (number of aircraft) and factor 6 (aircraft density) are the most important, reflecting that the traffic volume in the current period has a greater impact on complexity; for aircraft performance factors (7~9), factor 7 (variance of aircraft ground speed) and factor 8 (ratio of standard deviation of ground speed to mean value) are the most important, reflecting that the fluctuation of aircraft ground speed has a greater impact on complexity. For potential conflict factors (10~23), factor 14 (heading disorder), factor 15 (speed disorder), factor 16 (separation speed), factor 17 (convergence speed) and factor 23 (horizontal conflict) are the most important. *Figure 6(b)* further shows the metric matrix of the full matrix type. As for the elements on the diagonal, although the value of the weight coefficient is different from that of the diagonal matrix, the influencing factors with the larger coefficients of the two are relatively consistent; as for the off-diagonal elements, the weight coefficients of most elements are near 0, showing that most of the interactions between factors are not significant. In addition, the weight coefficients of some elements are small negative numbers. This phenomenon often occurs between factors with positive correlation, such as factor 1 (number of aircraft) and factor 6 (aircraft density), factor 14 (heading disorder)) and factor 15 (speed disorder) etc., whose purpose is to weaken the impact of simultaneous changes in the two factors on the distance metric.

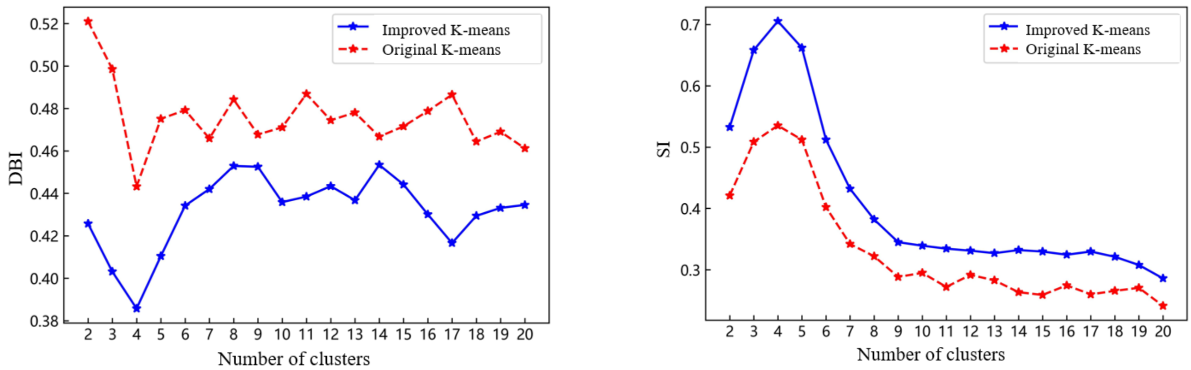


(a) *(b)*

Figure 6 – Distance metric matrix visualisation: a) Distance metric diagonal matrix; b) Distance metric full matrix

3.4 Analysis of complexity level classification

In order to determine the optimal number of clusters for complexity samples and evaluate the performance of different clustering algorithms, *Figure 7* shows the DBI and SI results of the K-means clustering algorithm based on Mahalanobis distance (i.e. improved K-means) and Euclidean distance (i.e. original K-means). In general, under the same number of clusters, the DBI of the improved K-means is smaller than that of the original K-means, while the SI index is the opposite. This phenomenon directly shows that the Mahalanobis distance based on metric learning can make the inside of the cluster more compact and the outside of the cluster more separated, effectively improving the performance of the clustering task. In addition, regardless of the clustering algorithm, when the number of clusters is 4, the two evaluation indicators achieve the optimal, so the complexity dataset should be divided into four levels, from low to high, namely, levels 1, 2, 3 and 4.



(a) *(b)*

Figure 7 – Clustering performance comparison: a) DBI; b) SI

Based on the analysis results in Section 3.3, factor 6 (aircraft density), factor 7 (aircraft ground speed variance), factor 14 (heading disorder), factor 17 (convergence speed) and factor 23 (horizontal conflict) are regarded as the key influencing factors of the complexity sample. *Table 3* further provides the changes in cluster centres at different levels. It can be found that as the values of these factors continue to increase, the level of air traffic complexity also gradually rises. In particular, for a new complexity sample, the distance metric matrix can be used to linearly transform the sample into the learned metric space. On this basis, by calculating its distance from the cluster centre of each complexity level, the level to which the complexity sample belongs can be obtained, thereby realising the complexity level assessment of online data.

Table 3 – Cluster centres of key factors of different complexity levels

Key factors	**Complexity level**			
	Level 1	**Level 2**	**Level 3**	**Level 4**
S_{dens}	7.64	9.40	16.60	20.05
σ_{gs}^2	5979.08	6359.71	6912.80	8721.20
S_{track_dis}	52.05	117.34	181.16	286.04
S_{conv}	169.74	584.67	912.39	1282.11
S_{inter_hori}	17.55	23.83	50.60	61.42

3.5 Analysis of complexity situation evolution

When the air traffic complexity level changes over time, by combining the cluster centres of key factors and the changes in the values of each factor, the key causes that affect the level changes can be inferred to a certain extent. Taking the complexity sample from 12:00–13:00 on 7 December 2019 as an example, *Figure 8* shows the change process of air traffic complexity level per minute. It can be seen that the complexity level changes continuously from 1 to 4 over time.

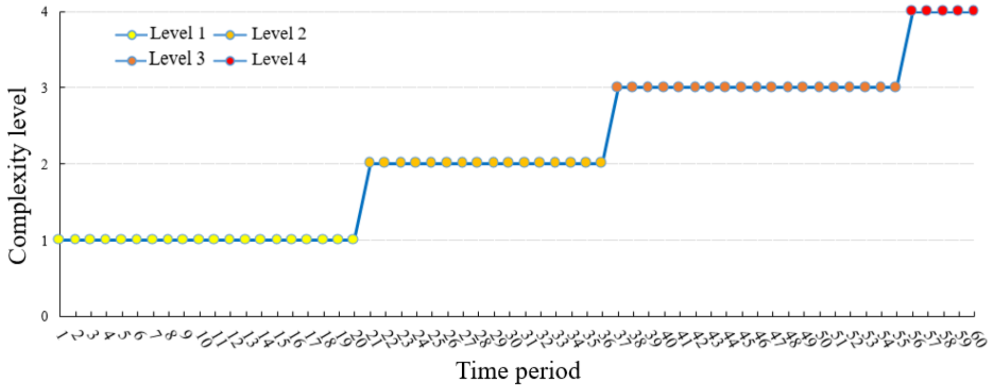


Figure 8 – Evolution of air traffic complexity levels over time

In order to identify the dominant factors that cause the evolution of the complexity level, based on the specific values of the key factors before and after the three changes, the gap in Mahalanobis distance between the factors in two adjacent periods is calculated, and the factors with a gap greater than 0.01 are displayed in bold, as shown in *Table 4*. It can be found that the dominant factors of changes at different levels are different, often consisting of 1–2 factors. Specifically, the dominant factor when the complexity level changes from 1 to 2 is aircraft density; the dominant factor when the complexity level changes from 2 to 3 is aircraft ground speed variance; and the dominant factor when the complexity level changes from 3 to 4 is heading disorder and horizontal conflicts.

Table 4 – Key factors before and after changes in complexity levels

Level changes	**Time period**	**Key factors**				
		S_{dens}	σ_{gs}^2	S_{track_dis}	S_{conv}	S_{inter_hori}
Level 1->2	12:19-12:20	8.05	5661.63	41.69	177.56	15
	12:20-12:21	11.75	5587.46	58.00	553.71	19
Normalisation factor gap		**0.038**	0.000	0.001	0.008	0.002
Level 2->3	12:35-12:36	13.81	6028.93	132.85	498.33	29
	12:36-12:37	14:20	7694.39	143.91	839.25	35
Normalisation factor gap		0.000	**0.011**	0.000	0.007	0.003
Level 3->4	12:54-12:55	19.44	7390.04	178.17	983.31	52
	12:55-12:56	18.65	8073.01	302.71	1117.54	67
Normalisation factor gap		0.002	0.002	**0.046**	0.001	**0.021**

4. CONCLUSION

Air traffic complexity analysis has always been the focus of air traffic researchers. Although controller subjective evaluation is considered a mainstream method, it still has certain limitations, such as inconsistent assessment of traffic situations. This paper attempts a new idea to solve this problem, which is to use highly reliable similarity information to replace level information that is difficult to accurately label. On this basis, distance metric learning and cluster analysis techniques are combined to analyse air traffic complexity, which can bring valuable insights from the aspects of factor importance, situation level classification, situation level evolution etc. A case study was conducted in a mid-southern area sector of China to illustrate the potential of the proposed framework.

Future work will verify the generality and effectiveness of the proposed method in more airspace sectors, and it is interesting to compare the differences in complexity analysis results for different airspace sectors or time periods. In addition, predicting the complexity of target sectors based on small-scale datasets of multiple adjacent sectors is another interesting topic.

ACKNOWLEDGEMENTS

This research was co-supported by the National Key R&D Program of China (2021YFB2601704), the Joint Funds of the National Natural Science Foundation of China (U2333207), the National Natural Science Foundation of China (52002178), the Sichuan Science and Technology Program (MZGC20240065) and the Fundamental Research Funds for the Central Universities (25CAFUC04053, 25CAFUC04062).

REFERENCES

[1] Gawade M, Zhang Y. Synthesis of remote air traffic control system and air traffic controllers' perceptions. *Transportation research record*. 2016;2600(1):49-60. DOI: 10.3141/2600-06.

[2] Antulov-Fantulin B, et al. Determining air traffic complexity–challenges and future development. *Promet-Traffic&Transportation*. 2020;32(4):475-485. DOI: 10.7307/ptt.v32i4.3401.

[3] Goranson J. Looking for trouble: How well the FAA's enhanced traffic management system predicts aircraft congestion. Massachusetts Institute of technology, 1993. URL: http://hdl.handle.net/1721.1/12155.

[4] Kopardekar P, Magyarits S. Measurement and prediction of dynamic density // Proceedings of the 5th USA/Europe Air Traffic Management R & D Seminar. 2003, 139. URL: https://www.academia.edu/68291479/Measurement_and_Prediction_of_Dynamic_Density.

[5] Laudeman IV, et al. Dynamic density: An air traffic management metric. NASA/TM-1998-112226, Moffett Field, CA: NASA 1998. URL: https://ntrs.nasa.gov/citations/19980210764.

[6] Gianazza D, Guittet K. Selection and evaluation of air traffic complexity metrics // 2006 ieee/aiaa 25TH Digital Avionics Systems Conference. IEEE. 2006:1-12. DOI: 10.1109/DASC.2006.313710.

[7] Gianazza D. Forecasting workload and airspace configuration with neural networks and tree search methods. *Artificial intelligence*. 2010;174(7-8):530-549. DOI: 10.1016/j.artint.2010.03.001.

[8] Zhu X, Cai K, Cao X. A semi-supervised learning method for air traffic complexity evaluation // 2017 Integrated Communications, Navigation and Surveillance Conference (ICNS). IEEE, 2017:1A3-1-1A3-11. DOI: 10.1109/ICNSURV.2017.8011885.

[9] Cao X, et al. A knowledge-transfer-based learning framework for airspace operation complexity evaluation. *Transportation Research Part C: Emerging Technologies*. 2018;95:61-81. DOI: 10.1016/j.trc.2018.07.008.

[10] Zhang W, et al. Sector complexity evaluation based on conditional generative adversarial networks. *Journal of Transportation Systems Engineering and Information Technology*. 2021;21(6):226-233. DOI: 10.16097/j.cnki.1009-6744.2021.06.026.

[11] Andraši P, et al. Subjective air traffic complexity estimation using artificial neural networks. *Promet-Traffic&Transportation*. 2019;31(4):377-386. DOI: 10.7307/ptt.v31i4.3018.

[12] Cao X, et al. A knowledge-transfer-based learning framework for airspace operation complexity evaluation. *Transportation Research Part C: Emerging Technologies*. 2018;95:61-81. DOI: 10.1016/j.trc.2018.07.008.

[13] Li B, et al. A deep unsupervised learning approach for airspace complexity evaluation. *IEEE Transactions on Intelligent Transportation Systems*. 2021;23(8):11739-11751. DOI: 10.1109/TITS.2021.3106779.

[14] Antulov-Fantulin B. Air traffic complexity model based on air traffic controller tasks. University of Zagreb. Faculty of Transport and Traffic Sciences, 2020. URL: https://repozitorij.fpz.unizg.hr/islandora/object/fpz:2296.

[15] Jurinić T, et al. Defining terminal airspace air traffic complexity indicators based on air traffic controller tasks. *Aerospace*. 2024;11(5):367. DOI: 10.3390/aerospace11050367.

[16] Pérez Moreno F, et al. Determination of air traffic complexity most influential parameters based on machine learning models. *Symmetry*. 2022;14(12):2629. DOI: 10.3390/sym14122629.

[17] Moreno FP, et al. Prediction of air traffic complexity through a dynamic complexity indicator and machine learning models. *Journal of Air Transport Management*. 2024;119:102632. DOI: 10.1016/j.jairtraman.2024.102632.

[18] Zhou ZH. A brief introduction to weakly supervised learning. *National science review*. 2018;5(1):44-53. DOI: 10.1093/nsr/nwx106.

[19] Yang L, Jin R. Distance metric learning: A comprehensive survey. Michigan State University. 2006;2(2):4. URL: https://www.cse.msu.edu/~rongjin/semisupervised/dist-metric-survey.pdf.

[20] Xing E, et al. Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems*. 2002:15. DOI: 10.5555/2968618.2968683.

[21] De Vazelhes W, et al. Metric-learn: Metric learning algorithms in Python. *The Journal of Machine Learning Research*. 2020;21(1):5447-5452. URL: http://jmlr.org/papers/v21/19-678.html.

[22] Pedregosa F, et al. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*. 2011;12:2825-2830. DOI: 10.5555/1953048.2078195.

张魏宁，潘卫军，杨昌其，朱新平，尹嘉男，杜婧涵

题目: 基于弱监督学习的主观空中交通复杂性分析

摘要
管制员主观评价是评估空中交通复杂性的重要手段之一。然而，人类专家的不一致性对复杂性分析模型的推理产生了负面影响。为了解决这一问题，本文提出以高可靠的交通态势相似度作为标注信息，构建弱监督的空中交通复杂性数据集。在此基础上，训练距离度量学习模型，生成满足相似关系的距离度量矩阵。最后，结合 K-means 算法，实现有偏好的复杂度态势等级分类和演变分析。以中国中南部区域扇区的实际运行数据为例，验证了所提方法的有效性。实验结果表明，飞机密度、飞机地面速度变化、航向紊乱、收敛速度和水平冲突对复杂度态势的影响较大。与基于欧氏距离的 K-means 算法相比，度量学习算法的最优轮廓系数和 Davidson-Boldin 指数分别提高了 31.80%和 12.97%。此外，还证实了态势演变是由一个或两个关键影响因素驱动的。

关键字
空中交通复杂性; 主观相似性; 弱监督; 距离计量学习; 聚类分析