Chinese Journal of Aeronautics, (2017), xxx(xx): xxx–xxx

Chinese Society of Aeronautics and Astronautics
& Beihang University

Chinese Journal of Aeronautics

cja@buaa.edu.cn
www.sciencedirect.com

Measuring air traffic complexity based on small samples

Xi ZHU, Xianbin CAO *, Kaiquan CAI

School of Electronics and Information Engineering, Beihang University, Beijing 100191, China
Beijing Key Laboratory for Network-Based Cooperative Air Traffic Management, Beijing 100083, China

Received 30 May 2016; revised 20 August 2016; accepted 30 November 2016

KEYWORDS

Air traffic control;
Air traffic complexity;
Correlation analysis;
Ensemble learning;
Feature selection

Abstract Air traffic complexity is an objective metric for evaluating the operational condition of the airspace. It has several applications, such as airspace design and traffic flow management. Therefore, identifying a reliable method to accurately measure traffic complexity is important. Considering that many factors correlate with traffic complexity in complicated nonlinear ways, researchers have proposed several complexity evaluation methods based on machine learning models which were trained with large samples. However, the high cost of sample collection usually results in limited training set. In this paper, an ensemble learning model is proposed for measuring air traffic complexity within a sector based on small samples. To exploit the classification information within each factor, multiple diverse factor subsets (FSSs) are generated under guidance from factor noise and independence analysis. Then, a base complexity evaluator is built corresponding to each FSS. The final complexity evaluation result is obtained by integrating all results from the base evaluators. Experimental studies using real-world air traffic operation data demonstrate the advantages of our model for small-sample-based traffic complexity evaluation over other state-of-the-art methods.

1. Introduction

Air traffic complexity is an objective and critical metric for evaluating the operational status of given airspace. From the perspective of complexity science, it can be summarized as three types of complexity contained in an air traffic management system (ATMS): the complexity embedded in the global operational pattern, the complexity contained in the relationships between various elements, and the complexity embodied in the uncertainty of the evolutionary trend.[1,2] Based on this definition, we can infer that air traffic complexity has a dominant influence on the workload of the air traffic controller (ATCo) because it brings the ATCo difficulty in perceiving traffic situations and making right decisions. In other words, ATCos are more likely to increase operational errors with higher traffic complexity.[3] Therefore, air traffic complexity is

* Corresponding author at: School of Electronics and Information Engineering, Beihang University, Beijing 100191, China.
E-mail address: xbcao@buaa.edu.cn (X. CAO).

Peer review under responsibility of Editorial Committee of CJA.

ELSEVIER | Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.cja.2017.04.018
1000-9361 © 2017 Chinese Society of Aeronautics and Astronautics. Production and hosting by Elsevier Ltd.

a critical factor that affects the operational safety in ATMS, and in turn ultimately limits airspace capacity.[4]

Today's ATMS, composed of numerous airspace sectors with varying air traffic flow, is a large-scale and rapidly evolving complex dynamical system, and thus air traffic complexity is consistently changing over time and sectors. For a specific sector, a mismatch of excessive air traffic complexity and limited traffic management ability frequently occurs, which may lead to airspace congestion and flight delays. To avoid such situations, we should effectively tune the operational status of the sectors by traffic management methods, such as air traffic flow management and dynamic airspace configuration, to balance the traffic complexity and the controllability of each sector. To implement these techniques, a reliable measurement of air traffic complexity is needed. Thus, air traffic complexity evaluation has become a popular research topic in the air traffic management (ATM) field.

To measure air traffic complexity, a direct approach is to define a tangible complexity indicator that can be explicitly formulated. Many scholars define the complexity indicator by a traffic attribute that is identified by them as the predominant representative for traffic complexity, such as the difficulty of potential conflict resolution,[5–7] the probability of conflict occurring,[8–10] and the disorder of traffic trajectories.[11,12] Note that each indicator of this type depicts air traffic complexity from a certain angle, which has limitations because the ATMS includes so many elements. For example, the former two indicators mentioned above cannot reflect traffic surveillance complexity. Thus, the definition perspectives of these indicators are insufficient for characterizing traffic complexity comprehensively.

There is another complexity measurement approach that has a more comprehensive view. Considering that air traffic complexity is the result of complicated interactions among a range of traffic attributes (complexity factors), many scholars use machine learning technique to measure complexity. Gianazza[13,14] and Xiao et al.[15] have respectively advanced two representative machine learning-based complexity evaluation models that achieved satisfactory performance through fully training on a large number of samples. Nevertheless, in the real world, a large sample set can be very difficult to obtain due to the expensive cost of accurately labeling the complexity value for the complexity sample (a complexity sample includes a collection of complexity factors and a corresponding complexity degree). The labeling work needs real-time participation of ATCos during the control task, which is time-consuming and labor-intensive. Therefore, in most cases, only a small number of samples are available for training the complexity evaluation model. In addition, the operational rules of ATMS are changing slowly, and the complexity generation laws are also evolving gradually. Thus, complexity samples and the evaluation model should be updated occasionally. For this reason, constructing large dataset and retraining evaluation model would be considerable burdens. Therefore, it is necessary to develop an improved model to accommodate the real-world applications with limited sample set.

In this paper, a novel machine learning method for rating sectors' air traffic complexity levels with small dataset is presented. In our work, air traffic complexity within a sector is classified into three levels: Low, Normal and High. Specifically, the low complexity level indicates a simple traffic pattern and a waste of control resources because the workload is much less than that the ATCo can provide. The normal complexity level indicates a balance between the traffic control demand and the ATCo's control ability. Therefore, at this level, the control resources can be effectively used while safety is ensured. The high complexity level means that the traffic is difficult to control, and the workload is high so that the ATCo is likely to increase operational errors. In this context, our model can be used as a decision support tool. The traffic complexity level output by our model can help ATCos make tactic control decisions, such as splitting or merging sectors. In consideration of the small sample set, we expect to obtain satisfactory evaluation results by encouraging mining of the classification information contained in every dimension of each sample. Hence, the first step of our approach is to generate multiple small-size factor subsets (FSSs) by sampling factors from the "factor pool" (FP, the original factor set). Then, corresponding to every FSS, a base classifier is trained. Next, we integrate the evaluation results of all of the base classifiers to obtain the final result. Within this ensemble learning scheme, each factor in the FP can be included in multiple small FSSs, and thus has many "chances" to be learned by numerous base evaluators. Note that our approach is an improved version of a popular ensemble learning model—random subspace (RS).[16] The improvement lies in how the FSSs are generated. In the traditional RS, each FSS is generated by randomly selecting factors, whereas in our approach, the factors' noise and independence analysis is referenced to generate more efficient and compact FSSs that include fewer noisy and redundant factors. Therefore, our FSS generation strategy further facilitates the expression of factors' information, and good complexity evaluation results should be achieved.

The remainder of this paper is organized as follows: Section 2 reviews representative air traffic complexity measurements proposed by predecessors; Section 3 elaborates the proposed ensemble learning model designed for rating sectors' air traffic complexity levels based on small samples; Section 4 presents the experimental studies and the analysis of the results; Section 5 concludes this paper and suggests future research work.

2. Related work

To date, numerous air traffic complexity evaluation methods have been proposed by many scholars and engineers. As mentioned in Section 1, these methods broadly fall into two categories.

The first category is to characterize air traffic complexity by an explicitly formulized indicator that describes the complexity from a certain angle. For example, the input-output approach[5–7], proposed by Lee et al., defines traffic complexity as "how difficult" a given traffic situation is in terms of the control activity required to resolve it in response to a change in "reference signal", that is, the presence of a new aircraft entering the airspace. Besides Lee, Prandini et al.[8–10] proposed a mid-term air traffic complexity characterization approach based on the occurrence probability calculation of multiple aircraft converging within a specific distance, and the aircraft future flying process is modeled as Brownian motion. Another representative metric, proposed by Delahaye et al., is based on the Lyapunov exponent (LE).[11,12] In this approach, the ATMS is modeled by nonlinear differential evolution equations, and