Transportation Research Part C 95 (2018) 280–294

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area[☆,☆☆]

Zhengyi Wang[*], Man Liang, Daniel Delahaye

Ecole National de l'Aviation Civile, 7, Avenue Edouard Belin, 31055 Toulouse, France

ARTICLE INFO

Keywords:
Air traffic management
4D trajectory prediction
Clustering
Multi-cells neural network
Machine learning
Data mining

ABSTRACT

4D trajectory prediction is the core element of future air transportation system, which is intended to improve the operational ability and the predictability of air traffic. In this paper, we introduce a novel hybrid model to address the short-term trajectory prediction problem in Terminal Manoeuvring Area (TMA) by application of machine learning methods. The proposed model consists of two parts: clustering-based preprocessing and Multi-Cells Neural Network (MCNN)-based prediction. Firstly, in the preprocessing part, after data cleaning, filtering and data re-sampling, we applied principal Component Analysis (PCA) to reduce the dimension of trajectory vector variable. Then, the trajectories are clustered into several patterns by clustering algorithm. Using nested cross validation, MCNN model is trained to find out the appropriate prediction model of Estimated Time of Arrival (ETA) for each individual cluster cell. Finally, the predicted ETA for each new flight is generated in different cluster cells classified by decision trees. To assess the performance of MCNN model, the Multiple Linear Regression (MLR) model is proposed as the comparison learning model, and K-means++ and DBSCAN are proposed as two comparison clustering models in preprocessing part. With real 4D trajectory data in Beijing TMA, experimental results demonstrate that our proposed model MCNN with DBSCAN in preprocessing is the most effective and robust hybrid machine learning model, both in trajectory clustering and short-term 4D trajectory prediction. In addition, it can make an accurate trajectory prediction in terms of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) with regards to comparison models.

1. Introduction

4D trajectory prediction is the core element of future air transportation system (FAA, 2010). During the development of Trajectory Based Operation (TBO) concepts of Single European Sky ATM Research (SESAR) and Next Generation Air Transportation System (NextGen) programs, to improve the predictability of air traffic is a crucial and fundamental research topic. 4D trajectory prediction refers to the calculation and prediction of longitude, latitude, altitude and time on the future waypoint sequence based on the historical flight data. According to the time scale, 4D trajectory prediction can be divided into two categories (Guan et al., 2014):

[☆] This article belongs to the Virtual Special Issue on "AI in ATM".
[☆☆] This paper is an extension of work originally reported in SESAR Innovation Days 2017 by the authors Wang et al. (2017).
[*] Corresponding author.
E-mail address: zhengyi.wang@alumni.enac.fr (Z. Wang).

https://doi.org/10.1016/j.trc.2018.07.019
Received 12 January 2018; Received in revised form 5 July 2018; Accepted 21 July 2018

1. Tactical (short-term) trajectory prediction: A prediction in a short period within several minutes or even shorter. Since the prediction scale is relatively small, minor change may have significant impact on prediction results. Therefore, tactical trajectory prediction requires as much information as possible. Flight-related information contained in radar or ADS-B data is usually taken.
2. Strategical (long-term) trajectory prediction: A kind of prediction before departure based on the flight plan, which provides the prediction from a macroscopic view. It is mostly applied to fuel consumption and airspace flow evaluation.

4D trajectory prediction can be influenced by numerous factors, such as aircraft weight, pilot actions, wind and temperature. These uncertainties will not only make it difficult to improve the prediction accuracy, but also will decrease the prediction process efficiency as the prediction time becomes longer (Tastambekov et al., 2014). In this paper, we propose a novel short-term trajectory prediction model, which combines different machine learning techniques, to address the problem of 4D trajectory prediction in Terminal Manoeuvring Area (TMA). The model is mainly divided into 2 parts: preprocessing model and prediction model. The preprocessing model extends and improves the trajectory clustering method proposed by Gariel et al. (2011), plays a role of preparing data with high quality for the prediction part. In addition, it supports an efficient and accurate prediction process. It contains the following steps: data cleaning, filtering, re-sampling, Principal Component Analysis (PCA), clustering and model training. In the prediction model, a novel technique named Multi-Cells Neural Networks (MCNN) is proposed. It can make parallel trajectory prediction for various shapes of traffic patterns. With Neural Network (NN) for each prediction cell, an accurate prediction for the whole complex trajectories in TMA is realised. Thus, the main contributions of this paper are threefold:

1. A novel hybrid 4D trajectory prediction model combining clustering-based preprocessing and MCNN is developed. A general guideline for selecting the model structures is presented in detail.
2. The proposed model is robust with regards to the data source. The 4D trajectory in a specified airspace containing noise and abnormal trajectories can be effectively and efficiently processed by the preprocessing model, providing high-quality inputs to the prediction model.
3. The preprocessing step leads to a substantial increase in the accuracy of the prediction model. The prediction model based on NN has a much higher accuracy than the widely-used classical linear regression model in our 4D trajectory prediction task. A comparison study is conducted to demonstrate the effectiveness of the hybrid machine learning models. Comparisons are made on K-means++ and DBSCAN algorithms in the preprocessing part, Multiple Linear Regression (MLR) and NN model in the prediction part, as well as the combination of two parts.

2. Literature review

4D trajectory prediction has attracted numerous attention from worldwide researchers in the past decades. Considerable research efforts have been taken to enrich the 4D trajectory prediction approaches. From the view of methodology, the trajectory prediction models can be mainly classified into aircraft performance models and machine learning models (Hrastovec and Solina, 2016).

Aircraft performance models belong to physics-based approaches. The model structure is based on kinematic assumptions. The model parameters are determined based on a model of the aircraft performance, the planned flight routes, the predicted atmospheric condition, and the expected command and control strategies given by pilots or Flight Management System (FMS) (known as Aircraft Intent). The most precise aircraft performance model is Base of Aircraft Data (BADA), which provides increased levels of precision in aircraft performance parameters for modelling and simulation Nuic et al. (2010). A variety of researches based on BADA and Aircraft Intent has been conducted. For example, Xi et al. (2008) presented a classified ADS-B-based trajectory prediction algorithm. Based on the state estimation by Kalman filter and intent information captured by a pretreatment and probability method, the aircraft trajectory can be predicted with computation efficiency and fewer errors. Porretta et al. (2008) presented a novel aircraft performance model in consideration of the effect of wind, for aircraft lateral guidance and a new procedure for speed estimation. The model input includes navigation data and aircraft intent information, based on BADA set. Simulation results show that the model is suitable for reliable trajectory prediction. Kaneshige et al. (2014) described the implementation and evaluation of a motion-based trajectory prediction function, which can increase the resiliency and robustness of TBO. Based on the performance index, such as the fuel consumption, flight time, the algorithm computes the difference between with trajectory prediction and without trajectory prediction. Clearly, these aircraft performance models have made significant contributions to trajectory prediction, however, most of these models made ideal assumptions, rarely considering the actual constraints and human behaviour factors. The prediction in this method is conducted one by one, which cannot reflect the interaction between trajectories. In addition, once the data resources are limited or poorly supported, the prediction models will be ineffective for short-term trajectory prediction, not to mention large-scale trajectory prediction problems.

As a branch of Artificial Intelligence (AI), machine learning has been developed over 30 years. One of the major function of AI is learning from experience and make predictions. The trend of recent years shows that machine learning is commonly used in trajectory prediction domain. Compared with those aircraft performance models, machine learning models were built with weak assumptions or even without assumptions. In most cases, it shows better prediction performance. For example, Fablec and Alliot (1999) used Neural Networks (NN) to predict an aircraft trajectory in the vertical profile. Trained by real historical trajectory dataset, 2 different model structures were presented. In the first method, the inputs include current altitude, the remaining altitude to reach Request Flight Level (RFL) and n past vertical speeds. The output is the next speed. The inputs of the second structure consist of starting altitude and remaining altitude to reach the RFL. n first initial speeds is the output. Simulation result showed that NN model gives better results than classical prediction functions based on the model of aircraft. Leege et al. (2013) introduced Generalized