Artificial intelligence prediction of air traffic flow rate at the Hong Kong International Airport

Kai-kwong Hon\textsuperscript{1,2}

\textsuperscript{1}Hong Kong Observatory, Tsim Sha Tsui, Hong Kong, China
\textsuperscript{2}Corresponding author’s e-mail: kkhon@hko.gov.hk

Abstract. The Hong Kong Flight Information Region (HKFIR) is surrounded by the airspaces of Guangzhou, Sanya, Taipei and Manila. International/regional flight routes to and from the Hong Kong International Airport (HKIA) rank among the busiest in the world. Air traffic flow rate is a complex parameter influenced by a variety of factors such as weather, business operation and regulations, etc. Accurate prediction of flow rate allows airport management to optimise tactical planning and decision making, potentially benefitting airport operations and air traffic efficiency. Leveraging the concept of Big Data, this paper establishes a preliminary algorithm for artificial intelligence prediction of air traffic flow rate at HKIA through supervised machine learning. Results based on 2017 data showed positive short-term prediction skills as well as superior performance over persistence based on the previous day.

1. Introduction

The Hong Kong International Airport (HKIA) is one of the busiest international airports in the world, with total aircraft movements exceeding 400,000 in 2016 \cite{1}. The Hong Kong Flight Information Region (HKFIR), which borders the airspaces of Guangzhou, Sanya, Taipiei and Manila, contains many international and regional flight routes to and from HKIA which rank among the busiest in the world \cite{2}. Accurate prediction of flow rate allows airport management to optimise tactical planning and decision making, potentially benefitting airport operations and air traffic efficiency.

Air traffic flow rate is a complex parameter influenced by a variety of factors such as weather, business operation and regulations, etc. Such a complex, inter-disciplinary problem naturally lends itself to Big Data and machine learning methods \cite{3}. In fact, artificial intelligence techniques have been successfully applied in a number of aviation-related topics, ranging from prediction of turbulence \cite{4} to fuel consumption modelling \cite{5}. In this study, we tackle the problem of short-term prediction of air traffic flow rate, including both arrival and departure rate, at HKIA.

This paper is organised as follows. Section 2 describes the data and method used for establishing the machine learning prediction model. Section 3 describes the verification results. Brief discussions and conclusions are given in Section 4.

2. Data and method

In this section, we describe the data and method used in the establishment of the machine learning model for prediction of air traffic flow rate at HKIA.
2.1. Aircraft position data

HKO has installed an Automatic Dependent Surveillance – Broadcast (ADS-B) reception system since 2016 [6] for real-time reception of aircraft position data around Hong Kong. The reception antenna is located at the Tai Mo Shan Weather Radar Station (114.12°E, 22.41°N) which stands at over 900 m above sea level and provides unobstructed line-of-sight over its neighbouring airspace, with an effective reception range up to approximately 600 km (Figure 1).

![Figure 1. Sample display of the ADS-B aircraft position data as received by HKO. The white lines mark the HKFIR boundary.](image)

The ADS-B data is updated up to every second (or more) and contains various flight parameters such as callsign, latitudinal and longitudinal position, geometric height, flight level, ground speed and true air speed, etc. For the purpose of this study, only the positional data will be used. Quality control procedures are applied to filter out potentially erroneous data (e.g. missing or incomplete data fields, repetitive entries, discontinuous timestamps or trajectory, unphysical values in parameter space).

2.2. Air traffic data

Air traffic flow rate at HKIA may be derived from the actual number of aircraft movement, including arrivals and departures, at each hour. During the study period (and up to the time of writing), HKIA operates two parallel runways (the North and South Runways along the 070-250 direction) in independent segregated mode, supporting a maximum of a total of 68 movements per hour (Figure 2). In reality, the maximum number of arrival and departure flights per hour would be influenced by various operational and regulatory factors including weather [7] and aircraft wake turbulence separation [8].

![Figure 2. The existing two runways of HKIA, with the new Third Runway to their north under construction as of the time of writing. (Image taken from Google Earth.)](image)
The observed hourly arrival and departure rates will form the basis for training and validation of the machine learning prediction model.

2.3. Artificial intelligence algorithm

In this study, we apply the open-source machine learning algorithm of XGBoost [9] as implemented on the R software platform [10]. XGBoost, or “extreme gradient boosting”, is a computationally efficient artificial intelligence algorithm based on iterative ensembles of regression trees with gradient boosting, which is suitable for supervised learning involving large data sets.

Figure 3 shows the schematic diagram of the model development process. ADS-B and air traffic data from 2017 are used. The training period covers the 3-month period of April to June. The validation period covers July to November. The distribution of aircraft positions as derived from ADS-B data over the previous hour is used as the predictor, while the hourly arrival rate and departure rate over the next hour is used as the predictand. In this way, we generate short-term prediction of air traffic flow rate for HKIA at hourly intervals, each covering the next hour.

![Figure 3. Schematic diagram of the development of the machine learning prediction model.](image)

3. Verification Results

In this section, we present the validation results of the machine learning prediction model. Figure 4 compares the results of the machine learning prediction model against the corresponding actual arrival rate at HKIA. The x-axis shows the predicted values while the y-axis shows the observed values. The black line shows the reference slope of 1:1. In the diagram, the colour of each pixel indicates the number of data pairs falling into that category. The value represented by each colour is explained in the legend to the right.

It can be seen from Figure 4 that the machine learning prediction model results exhibit strong linear correlation with the observed values. The forecast-observation pairs generally clutter along the line of unity throughout the considered range of values. The regions with highest density of data points (including those in yellow, orange and red) also occur closest to the line of unity. Results for departure rate is shown in Figure 5. It can be seen that, similar to Figure 4, predictions from the machine learning model show good correlation with the observations across the whole range of observed values. These results demonstrate that the machine learning prediction model exhibits positive skill for both arrival and departure rate prediction at HKIA.

We may further examine the skill level of the machine learning prediction model by comparing its mean absolute error (MAE) against the “persistence” method. The “persistence” method refers to using the observed value from the same hour of the previous day as the predicted value for upcoming hour. As indicated in Figure 4 and Figure 5, the MAE of the machine learning prediction model are 1.9 and 2.4 respectively for arrival rate and departure rate. The performance of the “persistence” method is shown in Figure 6 and Figure 7 respectively for arrival rate and departure rate. By comparison, the “persistence”
method gives MAE of 2.5 for arrival rate and 3.2 for departure rate. This further confirms the skill level of the machine learning prediction model.

**Figure 4.** Density plot of the machine learning prediction algorithm results (x-axis) against the corresponding actual arrival rate (y-axis). The colour of each pixel indicates the number of data pairs in that category.

**Figure 5.** Same as Figure 4 except for departure rate.

**Figure 6.** Same as Figure 4 except using the “persistence” method.

**Figure 7.** Same as Figure 5 except using the “persistence” method.

4. **Discussions and Summary**

Air traffic flow rate is a complex inter-disciplinary parameter with considerable operational significance for airport management. This study presents one of the first efforts in applying artificial intelligence
techniques in making short-term predictions of this important parameter. By using aircraft position data around HKFIR with high spatio-temporal resolution, the XGBoost-based machine learning prediction model is able to produce skillful predictions for both arrival rate and departure rate over the next hour at HKIA during the study period. Furthermore, the machine learning prediction model also demonstrated superior skill level over the “persistence” method from the previous day in terms of MAE.

Despite the initially promising results, we acknowledge various limitations in the present study. Firstly, the data set should ideally be expanded to cover multiple years in order to better establish robustness of the prediction model. This may be achieved by collecting long-term data to extend the training and validation sets. We also bear in mind that the air traffic system, whether in Hong Kong or other locations around the world, are not static but rather constantly evolving. Regulatory requirements, aircraft performance and handling, as well as air traffic control procedures and practices may change from time to time, to name a few. As such, the prediction model might need periodic updating to keep up with the latest environment, which in turn creates additional challenges for long-term validation. Furthermore, weather is known to exhibit considerable impact on air traffic over a range of timescales. Methods need to be developed for effectively incorporating such effects in the prediction model, as with other non-linear influence such as time-lag or cascading effects well-known to the traffic engineering community [11]. These would be possible areas of focus for future studies.

References
[1] Hong Kong International Airport, 2017: https://www.hongkongairport.com/iwov-resources/file/the-airport/hkia-at-a-glance/facts-figures/2016e.pdf
[2] OAG, 2020: Busiest Routes 2020, http://www.oag.com
[3] K.K. Hon, 2018: Big data opportunities in high-impact public and aviation weather at the Hong Kong Observatory. WMO CBS TECO, 23-26 Mar 2018, Geneva, Switzerland
[4] Hon, K.K., C.W. Ng and P.W. Chan, 2020: Machine learning based multi-index prediction of aviation turbulence over the Asia-Pacific, Machine Learning with Applications, Volume 2, 2020, 100008, https://doi.org/10.1016/j.mlwa.2020.100008
[5] Uzun, M., M.U. Demirezen, G. Inalhan, 2021. Physics Guided Deep Learning for Data-Driven Aircraft Fuel Consumption Modeling, Aerospace, https://doi.org/10.3390/aerospace8020044
[6] Hong Kong Observatory, 2019: A new ADS-B ground station at Tai Mo Shan, https://www.hko.gov.hk/en/aviat/outreach/35th/adsb.htm
[7] Lui, G.N., R.P. Liem and K.K. Hon, 2020: Towards understanding the impact of convective weather on aircraft arrival traffic at the Hong Kong International Airport. IOP Conf. Ser.: Earth Environ. Sci. 569 012067, https://doi.org/10.1088/1755-1315/569/1/012067
[8] Hon, K.K. and P.W. Chan, 2017: Aircraft Wake Vortex Observations in Hong Kong. Journal of Radars, 2017, 6(6): 709-718, https://dx.doi.org/10.12000/JR17072
[9] Chen T, Guestrin C, 2016: XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM; p. 785–94
[10] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
[11] Zhai X Q, Ou C Q, Xing F, Huang H L, K.K.Hon, Tony Sze and Zhan Z Y, 2019: Hourly associations between weather factors and traffic crashes: non-linear and lag effects. Anal. Methods Accid. Res. https://doi.org/10.1016/j.amar.2019.100109