Developing Fight Delay Prediction Model using Machine Learning

Team Leader : K.Manishankar
Team Members :Anishvikram.B,Kiran Kannan,Aswin.S.

Literature Survey

Analysis of flight delay and causal factors is crucial in maintaining airspace efficiency and safety. However, delay samples are not independent since they always show a certain aggregation pattern. Therefore, this study develops a novel spatial analysis approach to explore the delay and causal factors which is able to take dependence and the possible problem involved including error correlation and variable lag effect of causal factors on delay into account. The study first explores the delay aggregation pattern by measuring and quantifying the spatial dependence of delay. The spatial error model (SEM) and spatial lag model (SLM) are then established to solve the error correlation and the variable lag effect, respectively. Results show that the SEM and SLM achieve better fit than ordinary least square (OLS) regression, which indicates the effectiveness of considering dependence by employing spatial analysis. Moreover, the outcomes suggest that, aside from the well-known weather and flow control factors, delay-reduction strategies also need to pay more attention to reducing the impact of delay at the previous airport.

Paper 1:
Author: H. Khaksar ,A. Sheikholeslami.
Year: 2019
Publisher: Scientia Iranica
Summary: These methods were tested on a U.S. flight dataset and then refined for a large Iranian airline network. Results showed that the parameters affecting delay in US networks are visibility, wind, and departure time, whereas those affecting delay in Iranian airline flights are fleet age and aircraft type. The proposed approaches exhibited an accuracy of more than 70% in calculating delay occurance and magnitude

in both the whole-network US and Iranian. It is hoped that the techniques put forward in this work will enable airline companies to accurately predict delays, improve flight planning, and prevent delay propagation.
References:
http://scientiairanica.sharif.edu/article_20020_0.html

Paper 2:
Author: Alice Sternberg, Jorge Soares, Diego Carvalho, Eduardo Ogasawara.
Summary: Commercial aviation is a complex distributed transportation system. It deals with valuable resources, demand fluctuations, and a sophisticated origin-destination matrix that need orchestration to provide smooth and safety operations. Furthermore, individual passenger follows her itineraries while airlines plan various schedules for aircrafts, pilots and flight attendants. Figure 1 illustrates a typical operation of a commercial flight. Stages can take place at terminal boundaries, airports, runways, and airspace, being susceptible to different kinds of delays. Some examples include mechanical problems, weather conditions, ground delays, air traffic control, runway queues and capacity constraints [103, 63, 3].
Reference:
https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00380-z

Paper 3:
Flight delay predictions and the study of its causal factors using machine learning algorithms.
Author: Cho Yin Yiu; Kam K. H. Ng; Kin Chung Kwok; Wing Tung Lee; Ho Tung Mo
Year: 2021
Publisher: IEEE - 2021 IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)
Summary: Several machine learning approaches were adopted in this research to predict flight delay, including the decision tree, random forest, k-nearest neighbour, Naïve Bayes, and artificial neural networks.

The results show that all algorithms achieved more than 80% of accuracy and artificial neural networks perform the best among the alternatives. While Naïve Bayes is the least accurate, k-nearest neighbour have the lowest F 1 score.
References:
https://ieeexplore.ieee.org/document/9633571

Paper 4:
A Deep Learning Approach for Flight Delay Prediction Through TimeEvolving Graphs
Author: Kaiquan Cai, Yue Li, YiPing Fang, Yanbo Zhu
Year:2021
Summary: Graph Convolutional Neural Network (GCN) Through extensive experiments, it has been shown that the proposed approach outperforms benchmark methods with a satisfying accuracy improvement at the cost of acceptable execution time. The obtained results reveal that deep learning approach based on graphstructured inputs have great potentials in the flight delay prediction problem.
Reference:
https://ieeexplore.ieee.org/document/9512525

Paper 5:
Identification, Characterization, prediction of Traffic flow patterns in multi airport systems.
Author: Mayara Condé Rocha Murça; Robert John Hansman
Year: 2019
Publisher: IEEE - IEEE Transactions on Intelligent Transportation Systems
Summary: The data-driven framework is based on a sequential application of machine learning methods on historical flight tracks, weather forecasts and airport operational data. A multi-layer clustering analysis is performed to mine spatial and temporal trends in flight trajectory data for traffic flow pattern identification. The results revealed significant variability in throughput and delay performance for different metroplex configurations, emphasizing the importance of anticipating the behavior of the metroplex as a system when forecasting individual airport

capacity. Future research goes along this direction by exploring the development of higher-fidelity models for airport capacity prediction that take as input detailed weather information and metroplex configuration forecasts in order to deliver probabilistic capacity forecasts for strategic TMI planning.
References: https://ieeexplore.ieee.org/document/8373742

Paper 6:
Flight Delay Prediction Based on Aviation Big Data and Machine Learning
Author: Guan Gui, Fan Liu, Jinlong Sun, Jie Yang, Ziqi Zhou, Dongxu Zhao
Year:2019
Summary: Long- Short Term Memory (LSTM) based method, Random Forest based model Experimental results show that long shortterm memory (LSTM) is capable of handling the obtained aviation sequence data, but overfitting problem occurs in our limited dataset. Compared with the previous schemes, the proposed random forestbased model can obtain higher prediction accuracy (90.2% for the binary classification) and can overcome the overfitting problem.
Reference: **https://ieeexplore.ieee.org/document/8903554**