LITERATURE SURVEY

Date	08-10-2022
Batch no	B2-2M4E
Project name	DEVELOPING A FLIGHT DELAY PREDICTION USING MACHINE LEARNING

Team Leader	Santhosh T
Team member 1	Pradeep surya R
Team member 2	Palanisamy V
Team member 3	Nishanth M.L

STUDY OF FLIGHT DEPATURE DELAY AND CASUAL FACTOR USING SPATIAL ANALYSIS.

AUTHORS: shaowu cheng , Ronald Wesonga

ABSTRACT: 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.

REFER LINK: https://www.hindawi.com/journals/jat/2019/3525912/

1. ANALYSIS OF AIRCRAFT ARRIVAL AND DEPARTURE DELAY CHARACTERISTICS

AUTHORS: Eric R. Mueller and Gano B. Chatterji

ABSTRACT: The increase in delays in the National Airspace System (NAS) has been the subject of several studies in recent years. These reports contain delay statistics over the entire NAS, along with some data specific to individual airports, however, a comprehensive characterization and comparison of the delay distributions is absent. Historical delay data for these airports are summarized. The various causal factors related to aircraft, airline operations, change of procedures and traffic volume are also discussed. Motivated by the desire to improve the accuracy of demand prediction in enroute sectors and at airports through probabilistic delay forecasting, this paper analyzes departure and arrival data for ten major airports in the United States that experience large volumes of traffic and significant delays. To enable such an analysis, several data fields for every aircraft departing from or arriving at these ten airports in a 21- day period were extracted from the Post Operations Evaluation Tool (POET) database. Distributions that show the probability of a certain delay time for a given aircraft were created. These delay-time probability density functions were modeled using Normal and Poisson distributions with the mean and standard deviations derived from the raw data. The models were then improved by adjusting the mean and standard deviation values via a least squares method designed to minimize the fit error between the raw distribution and the model. It is shown that departure delay is better modeled using a Poisson distribution, while the enroute and arrival delays fit the Normal distribution better. Finally, correlation between the number of departures, number of arrivals and departure delays is examined from a time-series modeling perspective.

REFERLINK:

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.9924&rep=rep1&type=pdf

2. MODELLING AIRPORT EFFICIENCY WITH DISTRIBUTIONS OF THE INEFFICIENT ERROR TERM: AN APPLICATION OF TIME SERIES DATA FOR AIRCRAFT DEPATURE DELAY

AUTHOR: Ronald Wesonga, Makerere University, School of Statistics and Planning, P.O. Box 7062, Kampala UGANDA

ABSTRACT: The study employs determinants of the aircraft departure delay to estimate airport efficiency. Two main parameters were applied to fit the stochastic frontier model using transcendental logarithmic function where both frontier and inefficiency models were generated. The estimated airport efficiencies over a period of 1827 days applying the half-normal and exponential distributions for the inefficiency error terms were (0.7498;?=0.1417, n=1827) and (0.8181;?=0.1224, n=1827) respectively. The correlation coefficient for the efficiency estimates (?=0.9791, n=1827, p<0.05) between the half-normal and exponential distributions showed no significant statistical difference.

Further analysis showed that airport inefficiency was significantly associated with higher number of persons on board, lower visibility level, lower air pressure tendency, higher wind speed and a higher proportion of arrival aircraft delays. The study offers a contribution towards assessing the dynamics for the distribution of inefficient error term to estimate airport efficiency by employing both meteorological and aviation parameters. The study recommends that although either half-normal or exponential distributions could be used; the exponential distribution for the error term was found more suitable when estimating the efficiency score for the airport.

REFERLINK:

https://gssrr.org/index.php/JournalOfBasicAndApplied/article/view/1229

3. FLIGHT DELAY CAUSALITY :MACHINE LEARNING TECHNIQUE IN CONJUCTION WITH RANDOM PARAMETER STATISTICAL ANALYSIS

AUTHORS: Seyedmirsajad Mokhtarimousav and Armin Mehrabi

ABSTRACT: The consequences of flight delay can significantly impact airports' on-time performance and airline operations, which have a strong positive correlation with passenger satisfaction. Thus, an accurate investigation of the variables that cause delays is of main importance in decision-making processes. Although statistical models have been traditionally used in flight delay analysis, the presence of unobserved heterogeneity in flight data has been less discussed. This study carried out an empirical analysis to investigate the potential unobserved heterogeneity and the impact of significant variables on flight delay using two modeling approaches. First, preliminary insight into potential significant variables was obtained through a random parameter logit model (also known as the mixed logit model). Then, a Support Vector Machines (SVM) model trained by the Artificial Bee Colony (ABC) algorithm, was employed to explore the non-linear relationship between flight delay outcomes and causal factors. The data-driven analysis was conducted using three-month flight arrival data from Miami International Airport (MIA). A variable impact analysis was also conducted considering the black-box characteristic of the SVM and compared to the effects of variables indented through the random parameter logit modeling framework. While a large unobserved heterogeneity was observed, the impacts of various explanatory variables were examined in terms of flight departure performance, geographical specification of the origin airport, day of month and day of week of the flight, cause of delay, and gate information. The comprehensive assessment of the contributing factors proposed in this study provides invaluable insights into flight delay modeling and analysis.

REFER LINK: https://www.sciencedirect.com/science/article/pii/S2046043022000119

4. A MULTILAYER PERCEPTRON NEURAL NETWORK WITH SELECTIVE-DATA TRAINING FOR FLIGHT ARRAIVAL DELAY PREDICTION

AUTHORS: Hajar Alla , Lahcen Moumoun and Youssef Balouk

ABSTRACT: Flight delay is the most common preoccupation of aviation stakeholders around the world. Airlines, which suffer from a monetary and customer loyalty loss, are the most affected. Various studies have attempted to analyze and solve flight delays using machine learning algorithms. This research aims to predict flights' arrival delay using Artificial Neural Network (ANN). We applied a MultiLayer Perceptron (MLP) to train and test our data. Two approaches have been adopted in our work. In the first one, we used historical flight data extracted from Bureau of Transportation Statistics (BTS). The second approach improves the efficiency of the model by applying selective-data training. It consists of selecting only most relevant instances from the training dataset which are delayed flights. According to BTS, a flight whose difference between scheduled and actual arrival times is 15 minutes or greater is considered delayed. Departure delays and flight distance proved to be very contributive to flight delays. An adjusted and optimized hyperparameters using grid search technique helped us choose the right architecture of the network and have a better accuracy and less error than the existing literature. The results of both traditional and selective training were compared. The efficiency and time complexity of the second method are compared against those of the traditional training procedure. The neural network MLP was able to predict flight arrival delay with a coefficient of determination of 0.9048, and the selective procedure achieved a time saving and a better score of 0.9560. To enhance the reliability of the proposed method, the performance of the MLP was compared with that of Gradient Boosting (GB) and Decision Trees (DT). The result is that the MLP outperformed all existing benchmark methods.

REFER LINK: https://www.hindawi.com/journals/sp/2021/5558918/

5. EMPIRICAL ANALYSIS OF AIRPORT NETWORK AND CRITICAL AIRPORTS

AUTHORS: Cong ,Wei, Hu, Minhua, Dong bin, Wang Yanjun , feng ,cheg

ABSTRACT: Air transport network, or airport network, is a complex network involving numerous airports. Effective management of the air transport system requires an in-depth understanding of the roles of airports in the network. Whereas knowledge on air transport network properties has been improved greatly, methods to find critical airports in the network are still lacking. In this paper, we present methods to investigate network properties and to identify critical airports in the network. A novel network model is proposed with airports as nodes and the correlations between traffic flow of airports as edges. Spectral clustering algorithm is developed to classify airports. Spatial distribution characteristics and infraclass correlation of different categories of airports are carefully analyzed. The analyses based on the fluctuation trend of distance-correlation and power spectrum of time series are performed to examine the self-organized criticality of the network. The results indicate that there is one category of airports which dominates the self-organized critical state of the network. Six airports in this category are found to be the most important ones in the Chinese air transport network. The flights delay occurred in these six airports can propagate to the other airports, having huge impact on the operation characteristics of the entire network. The methods proposed here taking traffic dynamics into account are capable of identifying critical airports in the whole air transport network.

REFER LINK: https://www.sciencedirect.com/science/article/pii/S1000936116000376

6. ESTIMATING THE FLIGHT DEPATURE DELAY DISTRIBUTIONS – A STATISTICAL APPROACH WITH LONG- TERM TREND AND SHORT TERM PATTERN

AUTHORS: Yufeng Tu ,Michel O Ball & Wolfgang S Jank

ABSTRACT: In this article we develop a model for estimating flight departure delay distributions required by air traffic congestion prediction models. We identify and study major factors that influence flight departure delays, and develop a strategic departure delay prediction model. This model employs nonparametric methods for daily and seasonal trends. In addition, the model uses a mixture distribution to estimate the residual errors. To overcome problems with local optima in the mixture distribution, we develop a global optimization version of the expectation–maximization algorithm, borrowing ideas from genetic algorithms. The model demonstrates reasonable goodness of fit, robustness to the choice of the model parameters, and good predictive capabilities. We use flight data from United Airlines and Denver International Airport from the years 2000/2001 to train and validate our model.

REFERLINK: https://www.tandfonline.com/doi/abs/10.1198/016214507000000257

7. MODELLING CONGESTION PROPAGATION IN MULTISTAGE SCHEDULE WITHIN AN AIRPORT NETWORK

AUTHORS: Xiaoxu Dai, Minghua Hu, Wen Tian, Hao Liu

ABSTRACT: In order to alleviate flight delay it is important to understand how air traffic congestion evolves or propagates. In this context, this paper focusses on the aggravation of airport congestion by the accumulation of delayed departure flights. We start by applying a heterogeneous network model that takes congestion connection/degree into consideration to predict departure congestion clusters. This is on the basis of the fact that, from a micro perspective, the connection between congestion and discrete clusters can be embodied in models. However, the results show prediction to be of high accuracy and time consuming due to the complexities in capturing the connection in congested flights. The problem of being highly time consuming is resolved in this paper by improving the models by stages. Stage partitioning based on the variation of delay clusters is similar to the typical infectious cycle. For heterogeneous networks the model can describe the congestion propagation and its causes at the different stages of operation. If the connection between flights is homogeneous, the model can describe a more indicative process or trend of congestion propagation. In particular, for single source congestion, the simplified multistage models enable short-term prediction to be fast. Furthermore, for the controllers, the accuracy of prediction using simplified models can be acceptable and the speed on the prediction is significantly increased. The simplified models can help controllers to understand congestion propagation characteristics at different stages of operation, make a fast and short-term prediction of congestion clusters, and facilitate the formulation of traffic control strategies.

REFERLINK:

https://www.researchgate.net/publication/327056786_Modeling_Congestion_Propagation_in_Multistage_Schedule_within_an_Airport_Network