International Journal of Computer Science and Engineering Communications
Volume.6, Issue.1 (2018): Page.1788-1793
www.ijcsec.com

Prediction of Airline Delays– A Survey

Priyadurga.S[1], Srimathi.K[2], Srinithi.N[3], Kirubha.M[4]
[1,2,3]UG Scholar, [4]Assistant Professor
Department of Computer Science and Engineering,
Sri Ramakrishna Institute of Technology, India
[4]m.kirubhacse@gmail.com

Abstract: *Airline delays cause the most economic impact on both airlines and passengers. There are many factors which affect the airline delays such as inclement weather conditions, miscommunications, check-in delays, congestion in air traffic, fueling, security issues, mechanical problems etc. The most common cause of airline delays is due to irregular signal delivery from source or destination and this can be predicted by using data mining algorithms. India's domestic flight data was extracted and used to train the model. To balance the unstable training data, sampling techniques such as SMOTE is applied. The k-Nearest-Neighbors is applied to build the models which can predict the flight delays and also the nearest location for landing of the aircraft safely.*

Keywords: *improved KNN Algorithm, decision tree, Naïve-Bayes, Decision Trees, Airline delay Prediction System.*

Citation: Priyadurga.S, Srimathi.K, Srinithi.N, and Kirubha.M. (2018). Prediction of Airline Delays– A Survey. *International Journal of Computer Science and Engineering Communications*, 6(1), 1788-1793. Article ID 6117881793.

Copyright © 2018 Priyadurga.S et al., This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION:

Aviation is the art or science of flying airplanes. To design production and maintenance of aircrafts. The development and operation of heavier than air crafts including airplanes and piloted or guided rocket ships. In aeronautical applications, satellite navigation has long been an additional means for localization purposes. The development of GNSS has provided a supplementary positioning service for many flight phases, in leisure flying as well as commercial air transport. An augmentation system for assisting pilots from the initial phase from take off to landing inclement weather conditions, miscommunications, check-in delays, congestion in air traffic, fueling, security issues, mechanical problems etc will help

the higher level of safety that will be required to cope with the continuous increase in the number of flights.

Literature Survey:

Sun Choi et al. explains about the airline delays are caused by inclement weather conditions[1]. This can be predicted by using data mining and supervised machine learning algorithms. Data were extracted and used to train the model. To overcome the effects of imbalanced training data, the combination of SMOTE and random under sampling techniques are applied. Decision trees, random forest, the AdaBoost and the k-Nearest-Neighbors were implemented to build models which can predict delays of individual flights. Then, each of the algorithm prediction accuracy and the receiver operating characteristic curve were compared. Flight schedule and weather forecast were gathered and fed into the model. Using those data, the trained model perform binary classification to predict whether a scheduled flight will be delayed or n-time.

Justin G. Fuller et al. describes autopilot systems can be prohibitively costly due in large part to the expense of validation and verification required to certify them[2]. The Federal Aviation Administration (FAA), NASA, and the US Air Force have been working to develop alternative certification methods to reduce this cost. In particular, run-time assurance (RTA) methods have recently been gaining momentum as a potential avenue to achieve this goal. The NCA-RTA-MPR system provides a framework for a system that ensures proper behavior of an advanced controller at run-time, potentially reducing V&V costs. The framework submits to analysis as a hybrid control system that can be formulated based on a model of pilot behavior.

Martin Gilje Jaatun et.al, specifies that Cyber Security Incident Management (CSIM) is an emerging paradigm for the Aviation Industry and it is the major cause to find out the process involved in the aviation sector in major airports[3]. Up to date only limited research has given the correct requirements to develop a tangible solution for the emerging incidents. Deployments of such incidents are obtained from the correct record of the flight system in the airports. This can apply a series of Data Mining (DM) approaches (i.e.) paper leverages are a good practice to obtain the correct result. Cyber Incident Response Management (CIRM) for the aviation domain uses the experiences and results from the DM systems. For other critical issues CIRM and CSIM combined with DM is used to solve the problem.

S Meenakshi Sundaram et.al, notifies that Aviation Accidents results in millions of deaths injuries worldwide each year which is a major public health concern [4]. It mainly leads to the cause of death and sever injuries. So to avoid or overcome this issue Data mining techniques can be applied.“Improved Oscillated Correlation Feature Selection (IOCFS)” and could be used as pre-checklist by the aviation agencies to avoid accidents and improve safety journey. We apply data mining technologies such feature selection and classification tasks to analyze recorded air accidents during a particular period of time based on years. After classification new data set- Aviation Accident Dataset (AAD) dealing aviation accidents and causes as factors for accidents is obtained. The top most cause of accidents is selected by our proposed novel feature selection method where by this method the causes can be easily obtained.

Maria Solntseva-Chaley et.al, specifies that Airspace Sectorization is necessary for accidents prevention irrespective of heavy workload prevailing in airport. A new Data mining technique, solving the problem, is described for Sectorization. Sectorization should be performed taking into account regular traffic of aircrafts. It allows fulfilling stable partition of aircraft intent trajectory sample into the groups, corresponding to the same runway approaches. This Method is taking into account of special geometric characteristics of multidimensional space trajectories of aircrafts. By this method it could be easily segregated and a perfect aviation system can be designed in airports and accidents can be avoided deucedly.

Richard M. Keller, specifies that the management of the aviation data system is quite complex in nature for the organizations such as government agency, airline, airframe manufacturer, or aviation service provider[6]. To overcome this nature of aviation system data models are developed to characterize and manage the data generated that are used and stored by a particular organization. Various data providers employ qualitatively different or varied data model. This varied data model encounters problems in exchanging data across organizational boundaries. So over the previous decades standardized data exchange models are used to address data interoperability. Ontology models came only in recent years from a confluence of research in the artificial intelligence, semantic web, and information science communities. Hereby we create new data model ontology for the data. By which some of the standardized data exchange model results are compared with a new type of data model based on ontologies. Based on these results aviation data management is done.

Xinmin Tang et.al, explains that 4D trajectory prediction is the core technology of new aviation system[7]. Data mining technique a Clustering Method combining Time Warp Edit Distance (TWED) with K-means algorithm is applied to improve the accuracy of nominal flight profile[5]. This process is worked out in two steps. First a series of historical trajectory data with same origin and destination are pre-processed sequentially to eliminate the outlier point. Second, a novel adaptive clustering algorithm is proposed in which the distance between different trajectories is calculated by TWED algorithm rather than the conventional elastic similarity measure. K clustering is used to determine density threshold under centers adaptively [20]. Based on matching rules of TWED the K nominal flight profiles in each cluster are fitted. Finally, the predicted trajectory containing various control intentions is used to forecast aircraft trajectory in advance to improve efficiency of airspace. The results inaccordance with accuracy and stability is verified based on the application of state-of-the-art clustering algorithm with dynamic time warping (DTW) and compared effectively.

Bryan Matthews et al. describes The process can identify operationally significant events due to environmental, mechanical, and human factors issues in the high dimensional Flight Operations Quality Assurance (FOQA).Data mining and Knowledge discovery process that can predict precursors to aviation safety incidents[8].This data contains information about the flight control systems, actuators, engines, landing gear, avionics, and pilot commands. A novel multivariate time series search algorithm is used to search for signatures of discovered anomalies on massive data sets. And also working to identify and incorporate other sources of data that corresponds to these flights such as radar track data and weather data, as these data will be useful in identifying anomalies. These models which help in model segmentation ad selecting the appropriate model for predicting user behavior.

David A. Pagels et al. describes the field of aviation is always searching for new ways to improve safety[9]. Due to the large amounts of aviation data collected daily, parsing through it all by hand would be impossible. To overcome these problems we have used data mining algorithms and they are Multiple Kernel Algorithms, Hidden Markov Models, Hidden Semi-Markov Models and Natural Language Processing. Supervised, Semi-Supervised and Unsupervised models are used to finding a function to describe data. Natural Language Processing focused on gathering meaningful data. New text classification approach causes in aviation reports with an emphasis on minority causes. To accomplish this, the Boot strapping algorithm was used to find causes based on key words contained in the aviation reports.

Nicolas Maille et al. specifies that Investigate contributions of clustering algorithms for flight safety management within airlines[10]. Based on independent analysis of a set of commercial flights with different methodologies, we identify the strengths of each approach. To design of an analysis process using a flight decomposition relying on crew activities and including discrete parameters that reflect crew actions. The result lay the foundations for a better use of such data-mining algorithms in the next generation of flight Data Monitoring used by airlines.

Yue Song et al. specifies that Trajectory prediction is an important technology for ensuring safety and efficiency of the air traffic. Hybrid estimation algorithm and intent interference algorithm are usually used to make long-term probabilistic trajectory prediction. Here data mining algorithms are used to process the historical radar data and to abstract a typical trajectory library. An improved trajectory prediction algorithm is proposed based on the typical trajectory, which is used as intent information to update the transition probability matrix, and is also used to propagate the nominal trajectory instead of the flight plan path. The prediction performance of the proposed algorithm is tested using real radar data from North China Air Traffic Management Bureau. The simulation results show that the improved algorithm has a better prediction performance and the prediction accuracy improved by 10% at most.

Douglas J.Equils et al. describes the Complexity of Flight Projects continues to grow, the management of information has become the underpinning of effective organizational communication[12]. Documentation is the preferred method for capturing and disseminating information in organizations and at the Jet Propulsion Laboratory. The goal of the documents is to capture the requirements, concepts, design ant the testing plan for a given project such that it is reviewable and can easily be shared across the project. The process for generating these documents on Flight project, however, is typically done through a series of informal inquires for older versions of a needed document which is then updated with project specific information. Document Quick Start improves on the process and ultimately how this stream-lined approach will reduce risks and costs to the next generation of flight projects at JPL.

Yang Hui et al. specifies the Complexity of maintaining time series, Similarity search have been used[13]. Time series is a large set of Flight data collected by FDR. Similarity search for Flight data is the basis to solve the large data set. Identifying Number of dimensions for flight data is too hard. To overcome these problem Discrete Fourier Transform (DFT) algorithms can be used. And also used map the original Flight data to the

frequency domain. The spatial data index structure is built using the Adaptive segmentation, and the Euclidean distance is used as the similarity measurement. The result shows that this algorithm has much higher time efficiency and accuracy rate rather than the sliding window regression algorithm.

Saybani Mahmud Reza et al. specifies that Noise is a big problem for people living near airports, therefore the public are looking for ways to reduce the noise[14]. Optimal solution would be flight paths that are farthest from those areas, and worst paths are those, that just go above them. There are two classes of paths, namely optimal and non-optimal ones. Neural network is used here, which is capable of recognizing patterns. They used some coordinates of various flight path as input for learning purpose of neural network and defined two classes representing the optimal and non-optimal flight paths. The results have shown that this technique is well capable of recognizing the optimal and non-optimal flight paths. This technique can be use to reduce the noise.

Mingqiang Chen et al. describes The large errors of traditional aerodynamic 4D trajectory prediction models, a prediction model based on concept tree multi-layer spatial association rule mining flight track data was proposed [15]. Firstly, the factors that had impact on total flying track were found out from flying history, and the flying track of next flight was predicted. Then from the flying position data, the positions where aircraft was located in each sampling period were worked out, and the whole prediction 4D trajectory was got.

CONCLUSION:

It is specified that different Data Mining strategies such as Adaboost, Random forest, Decision trees and k-Nearest-Neigh borswere applied to predict the airline delays accurately. The various other data sets and different types of algorithms can be considered for increasing the efficiency. In future working with k-Nearest-Neighbors algorithms give the prediction delay accurately.

REFERENCES

[1]Sun Choi, Young Jin Kim, Simon Briceno and Dimitri Mavris, 2016. "Prediction of Weather-induced Airline Delays Based on Machine Learning Algorithms"IEEE, PP.1-6.

[2] Justin G. Fuller, Loyd Hook, Nathan Hutchins, K. Niki Maleki, Mark A. Skoog,2016."Toward Run-Time Assurance in General Aviation and Unmanned Aircraft Vehicle Autopilots".

[3] Martin Gilje Jaatun, Rainer Koelle “Cyber Security Incident Management in the Aviation Domain” IEEE - 11th International Conference on Availability, Reliability and Security, 2016, pp.510-515.

[4] S Meenakshi Sundaram, Dr. S Senthamarai Kannan, Dr. S Appavu alias Balamurugan “Ensemble Feature Selector for Aviation Accident Analysis and Prevention” Asian Journal of Research in Social Sciences and Humanities, 2016, pp. 81-92.

[5] Maria Solntseva-Chaley "New data mining technique for multidimensional aircraft trajectories analysis" ITM Web of Conferences, 2016, pp.1-7.

[6] Richard M. Keller, "Ontologies for Aviation Data Management" Airspace Operations and Safety Program, 2015, pp.1-8.

[7] Xinmin Tang, Junwei Gu, Zhiyuan Shen, "A Flight Profile Clustering Method Combining TWED with K-means Algorithm for 4d Trajectory Prediction" 978-1-4799-8952-2/15©2015 IEEE, pp.1-9.

[8]Bryan Matthews,Santanu Das,Kanishka Bhaduri, Kamalika Das, Rodney MartinNikunj Oza, Ashok N. Srivastavaand John Stutz,2015. ″Discovering Anomalous Aviation Safety Events usingScalable Data Mining Algorithms"American Institute of Aeronautics and Astronautics, pp.1-12.

[9] David A. Pagels "Aviation Data Mining"University of Minnesota, Morris Morris, Minnesota, 2013, pp-1-7.

[10] Nicolas Maille."On the use of data-mining algorithms to improve FOQA tools for Airlines" 978-1-4673-1813-6/13©2013 IEEE, pp1-8.

[11]Yue Song, Peng Cheng, Chundi Mu. "An Improved Trajectory Prediction Algorithm based on Trajectory Data Mining for Air Traffic Management"978-1-4673-2237-9/12 ©2012 IEEE,pp.1-6.

[12] Douglas J. Equils."Methodologies for Improving Flight Project Information Capture, Storage and Dissemination" 978-1-4244-7351-9/11 ©2011 IEEE,pp.1-7.

[13] Yang Hui."Research on Similarity Mining for Flight Data″, Second International Workshop on Education Technology and Computer Science, 978-0-7695-3987-4/10 © 2010 IEEE, pp.1-4.

[14]Saybani Mahmoud Reza, Adel Lahsasna, Teh Ying Wah."Applied Data Mining Approach in Ubiquitous World of Air Transportation", Fourth International Conference on Computer Sciences and Convergence Information Technology,978-0-7695-3896-9/09 © 2009 IEEE, pp.1-5.

[15]Mingqiang Chen,"The Research of Flight Trajectory Based on Association Rule"Air Navigation Section, School of Flight Technology, Civil Aviation Flight University of China,978-1-4244-7255-0/11©2011 IEEE, pp.1-4.