IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 2, February 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2023 ||

| DOI:10.15680/IJIRSET.2023.1202091 |

Forecasting Flight Delays Using Clustered Models Based on Airport Networks

Dr. Mrunal K. Pathak[1], Siddhant Bhosle[2], Prabansh Maini[3], Prashant Gangthade[4], Nadiya Mujawar[5]

Dept. of Information Technology, AISSMS's Institute of Information Technology, Pune, Maharashtra, India

ABSTRACT: Estimating flight delays is important for airlines, airports and passengers because delays are one of the biggest costs of air travel. Each delay can cause further delay propagation. Thus, the delay pattern of an airport and the position of the airport in the network can provide useful information to other airports. We address the problem of forecasting delays for airports using network data and the delay patterns of the corresponding airports in the network. The proposed "Clustered Airport Modeling" (CAM) approach constructs representative time series for each airport cluster and fits a common model (eg, REG-ARIMA) to each airport using network-based features as regressors. The models are then applied to the data for each airport separately to predict flight delays at the airport. We also performed a network-based analysis of airports and found that the Betweenness Centrality (BC) score of is an effective feature for predicting flight delays. Experiments with flight data over seven years at 305 US airports show thatthe CAM provides accurate predictions of flight delays.

KEYWORDS: Stock Market Prediction, Supervised Machine Learning, Classification, Regression, Support Vector Machine (SVM), Artificial Neural Network (ANN.

I. INTRODUCTION

The main factors behind flight delays are technical problems with the aircraft, weather conditions, schedule conflicts, overuse of airport capacity and the distribution of delays among the,flights. Although these factors have traditionally been well studied, the patterns resulting from the structure of the airport network are still not well understood. In this paper, we investigate whether the location of an airport in the transport network and information about the corresponding airports improve the estimation of delay models. Our goal is to predict flight delays by incorporating airport network information and the similarity of airport delay patterns into the estimation models. An accurate forecast of flight delays is important both for optimizing airline operations and airport capacity planning.The airport network is first represented as a graphical structure where each airport is a hub and the number of flights between two airports is an edge nodes. A set of schedule-based features is extracted for airports. In particular, we match thescore midpoint, midpoint midpoint, articulation point, degree,and weighted measures to the context of aviation networks.

II. RELATED WORK

Jiadai Wang, Student Member, IEEE, Jiajia Liu, Senior Member, IEEE, and Nei Kato-In This paper surveys the networking and communication technologies in autonomous driving from two aspects: intra- and inter-vehicle. The intravehicle network as the basis of realizing autonomous driving connects the on-board electronic parts. The inter-vehicle network is the medium for interaction between vehicles and outside information

Guan Gui, Senior Member, IEEE, Fan Liu, Student Member, IEEE, Jinlong Sun.-This paper explores a broader scope of factors which may potentially influence the flight delay, and compares several machine learning-based models in designed generalized flight delay prediction tasks. To build a dataset for the proposed scheme, automatic dependent surveillancebroadcast (ADS-B) messages are received, pre-processed, and integrated with other information such as weather condition, flight schedule, and airport information.

Suvojit Manna, Sanket Biswas, RiyankaKunduSomnathRakshit, Priti Guptaand Subhas Barman-In this paper, the model has been implemented on the Passenger Flight on-time Performance data taken from U.S. Department of Transportation to predict the arrival and departure delays in flights. It shows better accuracy as compared to other methods.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2023 ||

| DOI:10.15680/IJIRSET.2023.1202091 |

HariBhaskarSankaranarayanan Amadeus Software Labs Bangalore, India-In this paper, we collected a dataset of on-time performance, 48 global airport flights and estimated queue time ofpassengers. In addition, we applied a logistic modeltree (LMT) machine learning method to predict passenger satisfaction based on factors such as airport on-time performance, number of flights, on-time arrangement, average delay and queuing time. Results are presented and discussed foradditional reviews and research articles.

Leonardo Moreira, ChristoferDantas, Leonardo Oliveira CEFET/RJ-This paper focuses on the unbalanced distribution of delay classes (presence and absence) by conducting an experimental evaluation of several pre-processingmethods to developmachine learning flight delay classification models. These models were created from a dataset that combines national aviation activities with meteorological conditions at airports.

Fengxiao Tang, Student Member, IEEE, Zubair Md. Fadlullah, Senior Member, IEEE, BominMao-This paper proposes SDN-IoT (Software Defined Network-based IoT) to improve transmission quality. In addition, the intelligent technique of deep learning has been widely explored in supercomputer SDN. Therefore, we first propose a new traffic load prediction algorithm based on deep learning to predict the future traffic load and network congestion.

Jinlong Sun, Student Member, IEEE, Zhilu Wu, Zhendong Yin, Member, IEEE, and Zhutian Yang-This paper proposed a learner-based information fusion algorithm for multiple sensor-based vehicle navigation solutions. Images of the reconstructed error covariance matrix were used to determine the information sharing coefficients during fusion of thesubsystems. The results show that the proposed SC-ELF algorithm outperforms existing methods in scenarios where sensors experience atypical observations and failures.

Nei Kato, Zubair Md. Fadlullah, Bomin Mao, Fengxiao Tang, Osamu Akashi, Takeru Inoue, and KimihiroMizutani-In this article, we explained why it is important to rethink how a heterogeneous network traffic control method, such as routing management, can be improved to handle massive traffic growth. In this sense, we envisioned that deep learning, which has recently emerged as a promising machine learning technique, can significantly improve the management of heterogeneous network traffic. A supervised deep learning system has been proposed that is trained on uniquely characterized inputs using traffic patterns observed in routers at the edge of the network being considered.

Yang Cong, Senior Member, IEEE, Ji Liu-In this paper, we propose a general model to solve the overabundance problem of online similarity learning from big data, which usually consists of two different redundancies: 1) feature redundancy, that is, it has redundant (irrelevant) features.training data; 2) list redundancy, ie.non-redundant (or relevant) objects are located in low-level space.

Shichao Zhang, Senior Member, IEEE, Xuelong Li, Fellow, IEEE, Ming Zong, Xiaofeng Zhu-This paper also proposes an improved version of thekTree method (i.e., the k*Tree method) that speeds up its testing phase by adding the information from training samples to kTree leaf nodes as training.samples located in leaf nodes, their kNNs, and the nearest neighbour of theirkNNs. We call the resulting decision tree the k * Tree, which allowskNN classification using a subset of the training samples from the leaf nodes, rather than all of the newly used training samples.

Table 1. Literature Survey

Sr no	Authors	Results	Challenges
1	Jiadai Wang, Student Member, IEEE, Jiajia Liu, Senior Member, IEEE, and Nei Kato	Veins (VEhicles In Network Simulation) is a simulator that combines the traffic simulator SUMO with the network simulator OMNeT++ and provides many modern features in programming, especially a full-featured WAVE model, leading to the more accurate results	The main control of traditional vehicles is in the hands of human beings. Human drivers are responsible for observing complex traffic environment, as well as operating vehicles directly according to their own judgement and traffic regulations
2	Guan Gui, Senior Member, IEEE, Fan Liu, Student Member, IEEE, Jinlong Sun	The proposed random forest-based model can obtain higher prediction accuracy (90.2% for the binary	comparing the predicted delay minutes and real delay minutes

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2023 ||

| DOI:10.15680/IJIRSET.2023.1202091 |

		classification) .	
3	Suvojit Manna, Sanket Biswas, RiyankaKunduSomnathRakshit, Priti Gupta and Subhas Barman	It shows better accuracy as compared to other methods. This model can be used to predict the delay in flights in various airports of United States of America accurately	Limited dataset used
4	Hari Bhaskar Sankaranarayanan Amadeus	we applied Logistic Model Trees (LMT) machine learning method for predicting the level of passenger satisfaction based on factors like an airport on time performance, the number of flights, on-time ranking, average delays and queue time. The results are presented and discussed for further insights and studies.	Connection times are not added due to the dependency on airline schedule at every hub. This may be a key factor for passenger satisfaction in case of inter-continental connections. This study doesn't consider passenger demographics, class of travel and other personal preferences associated with satisfaction levels.
5	Leonardo Moreira, ChristoferDantas, Leonardo Oliveira CEFET/RJ	Our results indicate the models that applied the balancing techniques performed much better in predicting the occurrence of delays, getting about 60% of hits	Small dataset used
6	Fengxiao Tang, Student Member, IEEE, Zubair Md. Fadlullah, Senior Member, IEEE, BominMao	The simulation result demonstrates that our proposal significantly outperforms conventional channel assignment algorithms.	Simulation designing complicated.
7	Jinlong Sun, Student Member, IEEE, Zhilu Wu, Zhendong Yin, Member, IEEE, and Zhutian Yang	SVM-CNN-based Fusion Algorithm for Vehicle Navigation Considering Atypical Observations	Vehicle navigation dataset not available.

III. EXISTING SYSTEM

In this system, we used supervised machine learning to anticipate arrival delays by including knowledge about airport networks. Try to improve the models for predicting flight delays, the network position of the airports and their similarity in terms of delay time-series are explored.

IV. METHODOLOGY

1. Support Vector Machine

- Support Vector Machine (SVM) is used to classify the breast cell. SVM Support vectormachines are mainly two class classifiers, linear or non-linear class boundaries.
- The idea behind SVM is to form a hyper plane in between the data sets to express which class it belongs to.
- The task is to train the machine with known data and then SVM find the optimal hyper planewhich gives maximum distance to the nearest training data points of any class

We have k sub-spaces so that there are k classification results of sub-space, called CL_SS1, CL_SS2,..., CL_SSk. Thus the problem is how to integrate all of those results. The simple integrating way is to calculate the mean value:

$$CL = \frac{1}{k}\sum_{i=1}^{k} CL\_SS_i$$

Or weighted mean value:

$$CL = \frac{1}{k}\sum_{i=1}^{k} w_i CL\_SS_i$$

Where wi is the weight of classification result of subspace SSi, and satisfies:

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2023 ||

| DOI:10.15680/IJIRSET.2023.1202091 |

$$\sum_{i=1}^{R} w_i = 1$$

The centroid of a hand is calculated as follows:

$$\overline{X} = \frac{\sum_{i=0}^{k} xi}{k}, \overline{Y} = \frac{\sum_{i=0}^{k} yi}{k}$$

Where $(\overline{X}, \overline{Y})$ represents the centroid of the hand, xi and yi are x and y coordinates of the ith pixel in the hand region and k denotes the number of pixels that represent only the hand portion.

In the next step, the distance between the centroid and the finger tip was calculated. For distance, the following Euclidean distance was used:

$$\text{Distance} = \sqrt{(x2 - x1)^2 - (y2 - y1)^2}$$

Where (x1, x2) and (y1, y2) represent the two co-ordinate values.

2. Naïve Bays Classification

- Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier.
- The Naive Bayes algorithm is called “naive” because it makes the assumption that the occurrenceof a certain feature is independent of the occurrence of other features.

The basis of Naive Bayes algorithm is Bayes' theorem or alternatively known as Bayes' rule or Bayes' law. It gives us a method to calculate the conditional probability, i.e., the probability of an event based on previous knowledge available on the events. More formally, Bayes' Theorem is stated as the following equation:

$$P(A/B) = \frac{P(B/A)\ P(A)}{P(B)}$$

Let us understand the statement first and then we will look at the proof of the statement. The componentsof the above statement are:

P (A/B): Probability (conditional probability) of occurrence of event A given the event B is true

P(A) and P(B): Probabilities of the occurrence of event A and B respectively

P(B/A): Probability of the occurrence of event B given the event A is true

The terminology in the Bayesian method of probability (more commonly used) is as follows:

A is called the proposition and B is called the evidence.

P(A) is called the prior probability of proposition and P(B) is called the prior probability of evidence.

P (A/B) is called the posterior.

P (B/A) is the likelihood.

This sums the Bayes' theorem as

$$\text{Posterior} = \frac{(\text{Likelihood}).(\text{Proposition prior probability})}{\text{Evidence pripor probability}}$$

V. CONCLUSION

In this paper, we incorporated airport network data and used plot-based scores such as central point average (BC) and articulation score to predict arrival delays. Airport network locations and airport delay time series simulations are explored as possible parameters to complement flight delay forecasting models. We presented Clustered Airport Modeling (CAM) which uses a REG-ARIMA model enhanced with clustering results. The CAM approach involves clustering and modeling steps that use a network of airports. The network is used both for graph-based clustering of the airports and as an exploratory variable in the forecasting model. Our experiments show that CAM gives more accurate results than the basic model (SARIMA) which was applied to each airport separately. The BC score is found to be an effective regressor in clustered REG-ARIMAs. In a comparison between ISM and CAM LSTM (Long-Short Term Memory), which is generally considered the cutting edge of prediction, CAM was found to be as good as LSTM, both of which outperformed ISM. This finding suggests that using network structure
in similar forecasting problems is a promising direction to pursue.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.118| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 2, February 2023 ||

| DOI:10.15680/IJIRSET.2023.1202091 |

REFERENCES

[1] Networking and Communications in Autonomous Driving: A Survey Jiadai Wang, Student Member, IEEE, Jiajia Liu, Senior Member, IEEE, and Nei Kato, Fellow, IEEE
[2] Flight Delay Prediction Based on Aviation Big Data and Machine Learning Guan Gui, Senior Member, IEEE, Fan Liu, Student Member, IEEE, Jinlong Sun, Member, IEEE, Jie Yang, Member, IEEE, Ziqi Zhou, Student Member, IEEE, and Dongxu Zhao, Student Member, IEEE 2019
[3] A Statistical Approach to Predict Flight Delay Using Gradient Boosted Decision Tree Suvojit Manna, Sanket Biswas, RiyankaKunduSomnathRakshit, Priti Gupta and Subhas Barman Department of Computer Science and Engineering Jalpaiguri Government Engineering College, West Bengal, India 2017
[4] An Exploratory Analysis for Predicting Passenger Satisfaction at Global Hub Airports using Logistic Model Trees HariBhaskarSankaranarayanan Amadeus Software Labs Bangalore, India 2016
[5] On Evaluating Data Preprocessing Methods for Machine Learning Models for Flight Delays Leonardo Moreira, ChristoferDantas, Leonardo Oliveira CEFET/RJ.
[6] An Intelligent Traffic Load Prediction Based Adaptive Channel Assignment Algorithm in SDN-IoT: A Deep Learning Approach Fengxiao Tang, Student Member, IEEE, Zubair Md. Fadlullah, Senior Member, IEEE, Bomin Mao, Student Member, IEEE, and Nei Kato, Fellow, IEEE 2018
[7] SVM-CNN-based Fusion Algorithm for Vehicle Navigation Considering Atypical Observations Jinlong Sun, Student Member, IEEE, Zhilu Wu, Zhendong Yin, Member, IEEE, and Zhutian Yang, Senior Member, IEEE 2018
[8] The Deep Learning Vision for Heterogeneous Network Traffic Control: Proposal, Challenges, and Future PerspectiveNei Kato, Zubair Md. Fadlullah, Bomin Mao, Fengxiao Tang, Osamu Akashi, Takeru Inoue, and KimihiroMizutani IEEE
[9] Online Similarity Learning for Big Data with Overfitting Yang Cong, Senior Member, IEEE, Ji Liu, Baojie Fan, Peng Zeng, Haibin Yu and JieboLuo, Fellow, IEEE 2016
[10] Efficient kNN Classification With Different Numbers of Nearest NeighborsShichao Zhang, Senior Member, IEEE, Xuelong Li, Fellow, IEEE, Ming Zong, Xiaofeng Zhu*, and Ruili Wang

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details