Modeling Flight Delay and Cancelation

Zhijie Wang
Stanford Center for Professional Development
Stanford University
Fremont, CA
zhw47@stanford.edu

Abstract— **This work uses flight record from Department of Transportation to predict whether a flight will be delayed by more than 30 minutes (either departure delay or arrival delay). The goal is to develop a policy to help a human traveler to make travel decisions under uncertainty, thus features used for modeling are limited to those observable at airport dashboard and some contextual information.**

Keywords—Flight Delay, Random Forest, Bayesian Network, Hidden Markov Model.

I. INTRODUCTION

The idea of working on flight delays came from personal work experience, as I happen to be a IT solution consultant and travel very frequently. For frequent travelers, traveling burnout is a constant struggle and last-minute meetings are bound to happen. There are two objectives in business travel decisions, to arrive on-time and minimize travel stress. Arriving on time for meeting and relaxed can lead to better presentation and negotiation outcomes. The actual travel decisions are two steps, booking and adjustment.

From past experience, the early morning flights are less likely to be delayed, because it rules out the possibility of aircraft not available, flight crew not available and other factors. Sunday evening, Monday morning and Thursday/Friday evenings are busy time for business travelers, because it is the time for them to fly out to client site and fly back to home. Modeling flight delay probability will help in making the first decision of when to fly out.

When stuck at the airport, due to unfortunate delays, business travelers, often time with premier member status, can arrange for standby or itinerary change. The current flight is delayed does not imply the next flight will be delayed for sure. There are situations when the flight at T is delayed by several hours to T+n and depart after the next flight which is scheduled to depart at T+1, where n>1. Modeling the airport congestion status can facilitate such decision making. Airport information terminal is often not updated on time; thus, it is also a very noisy signal for delay status for following flights.

In order to model flight delays, the input is a list of observable features of the flight at the airport information display (airline, departure airport, etc.) and some contextual information (distance, day of the week, etc.). Random Forests and Bayesian Networks are used to model flight delay probability and Hidden Markov model is used to model departure airport conditions. The flight delay probably is treated as a binary classification problem, Delay vs Non-Delay. Flight delay is if one of the conditions is met: canceled, departure delay > 30 min, arrival delay > 30 min. This problem has been studied in several research works. Flight Delay and Cancelation data set is also publicly available on Kaggle with lots of analysis.

Airport traffic congestion status is a time series prediction problem with Hidden Markov model of 2 states, Congested and Non-Congested. This is important for making decisions on the fly when a flight is delayed. Each flight has its allocated queue time for runway take off, and due to safety concern, flight can't merge into the queue as easily as highway lane change. If an airport is busy or large amount of flight delayed already, it is hard to reschedule a delayed flight into the queue, normally done ad-hoc.

Based on prior knowledge, each airport has its own unique characteristics, due to its geolocation, weather and relationships to other airports. For modeling the congestion state of origin airport, the time series is constructed by selecting all flights departing in the same day, and arranging them sequentially, order by scheduled departure time. Based on the personal experience, the size of the airport or the amount of flight to the airport does not necessarily indicate high likelihood of flight delay.

II. RELATED WORK

Several student projects in CS 229 class and several research papers have tried to model the flight delay problem. Most of the school work leveraged rich feature set, such as terrorist events in the past week, weather data [1], [2], [3]. In reality, most of these features are not observable to travelers when booking the original trip, as weather forecast for several days out is not reliable.

Researchers from NASA and FAA have performed in-depth analysis of flight delay causes and identified seasonality, origin airport and hour of departure as factors contributing to flight delay [4], [5]. This shows that there is not a strong correlation between amount of flights and the airports likelihood for delay. Some have adopted a statistical approach to model the flight delay with a factor graph [6].

III. DATASETS AND FEATURES

The data includes every single domestic flight from January 2013 to 2017 September, with scheduled time, actual time, origin airport, departure airport, delay time, flight number, airline, taken from Bureau of Transportation Statistics [7].

A. *Exploratory Data Analysis*

The average percentage of flight delay is 13.267% with standard deviation of 0.06481, and the average flight count is 81016 with standard deviation of 199787.37. Table I select top 10 airports by departure flights count and the likelihood a flight being delayed. These airports are also the subject of our model, as travelers are likely to travel to or transfer at one of these ten airports. There are about is about 27.3M flights total, and the top 10 airports include about 9.9 M flights. Training data and validation set is split via 80/20 ratio. Due to computation constrain, Bayesian Network Classifier is only trained with 9-month of data from 2017 and for Hidden Markov Model is trained with flights originated from Atlanta International Airport in 2017, which is about 1.7M flights.

TABLE I. TOP 10 BUSIEST AIRPORTS AND FLIGHT DELAY LIKELIHOOD

Origin	Percentage of Delay	Flight Count
ATL	12.06%	1789222
ORD	18.82%	1311383
DFW	15.97%	1130263
DEN	15.64%	1130263
LAX	14.33%	1019863
SFO	16.56%	782222
PHX	11.48%	767200
IAH	14.21%	735119
LAS	14.26%	679796
MSP	11.30%	599355

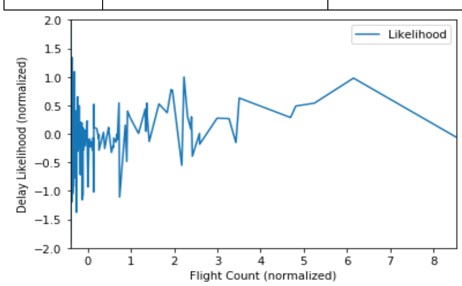

Fig. 1. Delay Likelihood vs Flight Count

B. Sequence Analysis on Consecutive Flight Delays

Fig. 2 shows the airport delay sequence length distribution. It is highly skewed to the right tail, which represents rare but severe weather conditions causing hundreds of flights to be canceled. The median length of consecutive delay is 64, and single flight delay accounts for 80.21% of total sequences.

The delay sequence is sorted by planned departure time for the flight, not by the actual departure time. It is possible that a long series of departure delay is broken up into 2 shorter sequences by one single flight that departs on time. This can happen when the flight control tower decides to allow some flights to depart according to original plan and reschedule delayed flights in ad-hoc fashion.

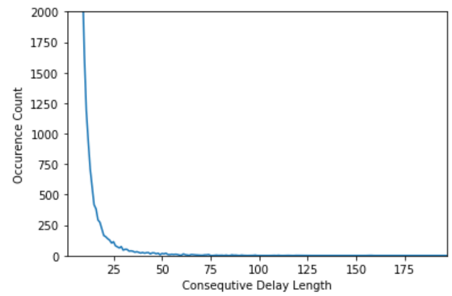

Fig. 2. Consecutive Delay Length Frequency

C. Feature Selection and Discretization

Several data points are continuous variables, such as departure time, air time, and distance. All time entries are stored as float number, i.e. 1846 means 18:46, The hour portion of the departure time is kept, discretize the flight time to hours, and discretize the distance by factor of 250. As stated before, the features are limited to observable ones at the airport: Month, Day of Month, Day of Week, Unique Carrier, Flight Number, Origin Airport ID, Destination Airport ID, Air Time, Distance, Computer Reservation System Departure Time and Delay. The Computer Reservation System Departure Time is the original scheduled departure time for flight to leave the boarding gates. Many features, such as Wheels Off Time, are removed.

D. Library Selection

Common python libraries such as, Pandas, Numpy and Scipy are used to preprocess data, calculate linear algebra and compute metrics. To train Random Forrest classifier, Scikit-learn's implementation is used

For Bayesian Network and Hidden Markov Models, several packages are explored: Bayespy, PyOpenPNL, Pomegranate, hmmlearn (a spin off from Scikit-learn) and Pgmpy. Most of these implementations are not up to date or not performing well. When training with Pgmpy, it caused Linux Kernel to kill the process, due to Out of Memory on a workstation with 32GB memory and 256 SSD SWAP space. Among all these libraries, Pomegranate package is the most stable one with support for GPU. Still some modification was made to the Pomegranate package on local machine to fix data serialization problem.

IV. FLIGHT DELAY CLASSIFICATION

A. Dummy Base Line

2 dummy classifiers (D1, D2) are created to be the baseline for comparison. D1 randomly predicts the flight delay with 50/50 split. D2 randomly predicts the flight as Delay with 13.267% (likelihood based on sample). The dummy classifier's performance is recorded in Table II.

TABLE II. DUMMY CLASSIFIER PERFORMANCE

Classifier	F1	Precision	Recall
D1	0.2128	0.5007	0.1351
D2	0.2335	0.8672	0.1349

B. *Random Forest Decision Tree Classifier*

Because flight delay only consists of about 13% of the datasets, the dataset is augmented to limit the overrepresented Non-Delay class. Without the limiting the Non-Delay class, the F1 score is merely 9% with Depth=25, the model performance stalled with additional depth.

With balanced training data set, Delay to Non-Delay as 1:4, the model achieved a F1 score of 23.78%, precision 100% but recall of 13.49% with Depth=10. The model performance stalled with additional depth.

The model performance is very low, compared to F1 score 82% achieve in [4]. After factoring the number of features (10 features vs 37 features from [4]), the model's performance is fair and more realistic. The Random Forest Decision Tree Classifier did marginally better than both D1 and D2, in terms of F1 and precision.

C. *Bayesian Network*

1) *Structure Learning*

To model we have some prior knowledge. Other work has tried to identify the causes for flight delay and produced Fig 3. [4]. Most of these features are not observable to travelers or any planning agent.

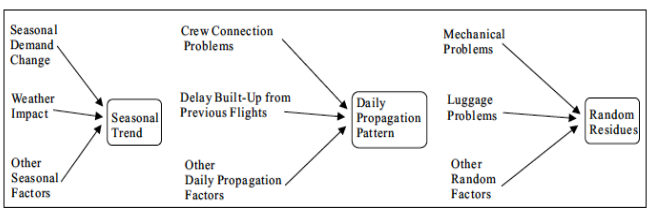

Fig. 3. Factors Influencing Departure Delay

Fig. 4 is manually constructed based on following believes:

a) Airline – different airlines will have different likelihood of flight delay. Some major airlines, to appeal to business travelers and improve their on-time rate, have spare aircraft available to handle issues

b) Airport – major airports, such as JFK, would have a higher chance of flight delay, due to the amount of traffic. Also, JFK is a prime example of crowded airspace. Within 37 miles of driving (actual point to point distance will be shorter), there are 3 airports, John F. Kennedy International Airport (JFK), LaGuardia Airport (LGA) and Newark Liberty International Airport (EWR).

c) Date – Travel date can imply holiday season or severe weather conditions in winter. Flight is more likely to be delayed in winter at Chicago O'Hare airport due to snow and high traffic at Christmas time.

d) Departure Time -- the effect of delay from previous flights will be minimized in the early morning.

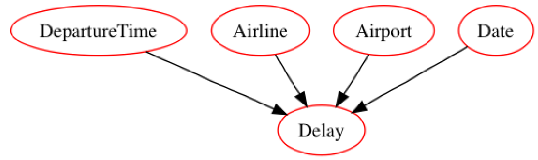

Fig. 4. Bayesian Network Structure Based on Human Knowledge

The Bayesian Network structure learning is very time consuming, a chow-liu tree [8] search strategy is used, the implementation is available in Pomegranate package. The algorithm was given the same feature set used in Random Forest classifier to perform structure learning and the result is this network, Fig. 5

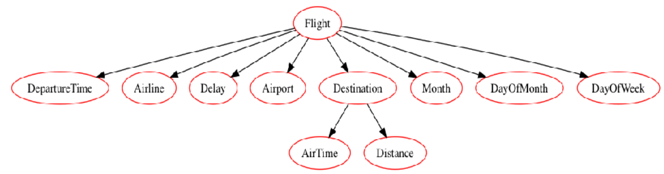

Fig. 5. Bayesian Network Constructed From Structure Learning

The model apparently failed to pick up the seasonality of flight delay as demonstrated in [6], shown in Fig 6. The spline is fitted on flight data from 2000 Denver airport data. This could because the training data set for Bayesian Network does not include enough seasonality, (Jan 2013 – Sept 2017, 57 months) or the seasonality is specific to certain airport, as Denver Airport ranked 5 in top 10 most weather-delayed airports [9]

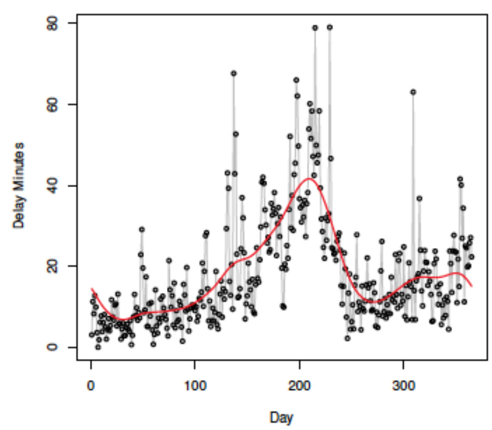

Fig. 6. Estimating the Seasonal Trend: A fitted smoothing spline that represents the seasonal trend from [4]

The final network structure after merging the two graphs is Fig 7.

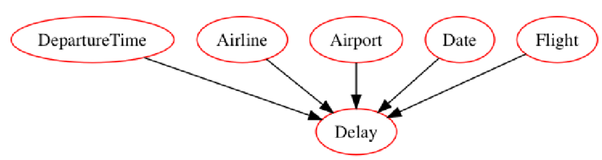

Fig. 7. Bayesian Network Merged

2) Parameter Learning

The model after structure learning, the model is trained with full data set to update the parameters with Maximum Likelihood Estimation algorithm. To reduce the parameter fitting time, each feature is discretized and modeled with Discrete Distribution. For fitting, pseudo count of 1 is added emission of each distribution, to prevent 0 probability for observations that are not included in training set.

The model performed very poor with F1 score of 0, as the model keep predicting Non-Delay for all data in validation set, due to skewness data set. The training time for Bayesian Network is reasonable, but it takes a very long time (> 30 min) to predict 100 samples on a MacBook Pro 2016 laptop. The exact cause of this performance is unknown (the likely suspect is Discrete Distribution calculation for each node).

V. Modeling Airport Congestion

As noted in other research, previous flight delays and other environmental conditions, such as weather and air traffic control can contribute to subsequent flight delay [1], [4]. In this work, the origin airport congestion states is modeled in two states, Congested and Non-Congested. The Hidden Markov Model is chosen, as the true state of the airport is not observable and the flight delay sequence is only a noisy indicator.

A. States Transition and Initial Belief

The initial state transition belief is based on human input. The sequence starts from None-start, and since it is the beginning of the sequence, it is likely to transit to Non-Congested (p=0.9) and in rare cases transit to Congested (0.1). If the airport is already busy and in Congested state, the delay will impact subsequent flights. For Congested state, it will likely to stay (p=0.9) and eventually clear out the delay queue (p=0.1), either by cancelation or rescheduling. If the airport is in Non-Congested, it is unsure whether it will become congested or not next, so $p(S_{t+1} = S_t | S_t) = 0.5$. The state transition network is shown in Fig 8. The model is then fitted with data to update the transition probabilities.

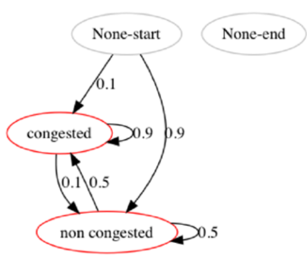

Fig. 8. Initial Hidden Markov Model with State Transition Probability Constructed From Human Knowledge

Fitting the data is an Expectation Maximization procedure. The true transition probabilities between hidden sates are unknown, Viterbi algorithm is not a good fit. The Baum-Welch algorithm, also known as Forward-Backward algorithm, is used. Given only the observable states, Baum-Welch algorithm produces most likely hidden transition probabilities and the most likely set of emission probabilities.

The only observable feature passed in for fitting is the flight delay time sequence and the time interval between each observation is not uniform. The Atlanta Airport is selected for modeling, as it has the most flights. The original training dataset is split into 1734 different time series, each represents one single day. The first entry in the sequence represents the delay status of first flight scheduled to depart and the final observation is the delay status of last flight scheduled to depart.

From Fig. 9 and Fig. 10, the Hidden Markov Model is able to detect large series of flight delay. The sequence used for Fig. 9 has 1133 flights, with 7.67% delay. The sequence used for Fig. 10 has 918 flights with 50% delay. In Fig. 10, the Hidden Markov Model properly captured the airport congestion transition, after there are more consecutive flight delays.

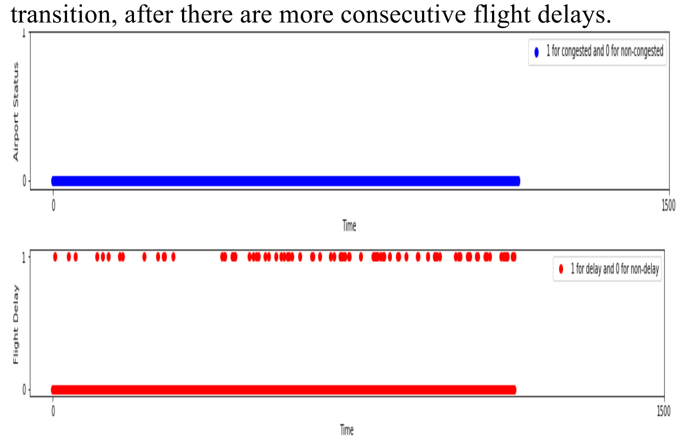

Fig. 9. A Sample of Hidden Markov Model Predicted State Sequence (blue) and Actual Flight Delay Sequence (red)

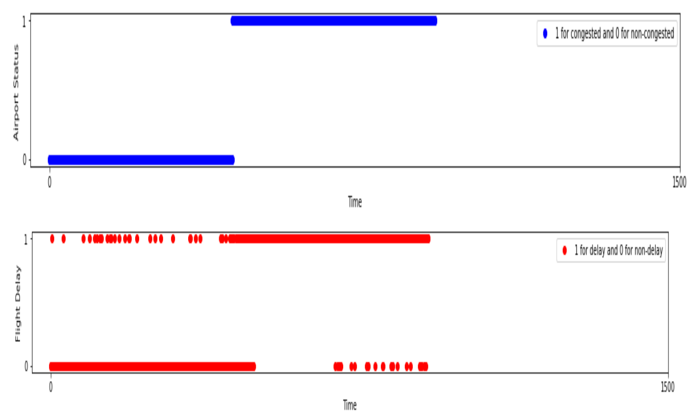

Fig. 10. Another Sample of Hidden Markov Model Predicted State Sequence (blue) and Actual Flight Delay Sequence (red).

The final hidden state transition model is plotted after the fitting the model with data, Fig 11. The state transition probability is similar to the initial belief, but more extreme. $p(S_{t+1,congested} | S_{t,congested}) = 0.99$ means that if an airport is in congested state it will stay in congested state and flights will be impacted by previous delay.

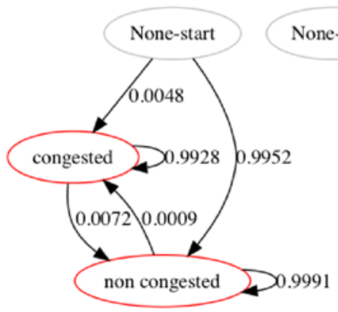

Fig. 11. Final Hidden Markov Model with State Transition Probability Fitted With Data

Conclusion And Future Work

This work applies Bayesian Network and Hidden Markov models, on two decisions stages when traveling by flight with limited features. See performance chart Table 3.

TABLE III. MODEL PERFORMANCE

Classifier	F1	Precision	Recall
D1	0.2128	0.5007	0.1351
D2	0.2335	0.8672	0.1349
Random Forest Classifier	0.2378	1.00	0.1349
Bayesian Network	0.0	0.0	0.0

And several lessons learned:

a) Models, such as Random Forest Classifiers, require lots of features and variety in data to train. From a pure modeling perspective, training data set can be enriched by engineering features from other data source. This strong dependency implies that all information must be available and accurate. Assuming a traveler is trying to make a booking decision 2 weeks ahead of time, the traveler can not predict the weather or terrorist attack incidents reasonably.

b) Bayesian Network models based on Maximu Liklihood Estimation are vulnerable to skewness in training dat. The MLE calculates expected liklihood based on sample observation.

Several areas can be further explored to improve the model performance:

c) Implement parallel computation backend for Bayesian Network. The current implementation in Pomegranate only leverages the GPU for faster matrix multiplication. The training could also be parallelized. This could significantly speed up the structure learning and parameter learning, and lead to faster model iteration and accommodate larger training dataset.

d) Incorporate other features. This work limits flight data strictly to the features observable at the airport information terminal and some contextual information. But there are additional features could be included, such as current weather condition at the departure airport.

e) Apply reinforcement learning. This works is focused on modeling states space and state uncertainty. It paves the way for next step reinforcement learning and derive a set of policies for travelers. The uncertainty of airport state makes it a Partial Markov Decision Problem.

References

[1] Duperier, Sauvestre and Leaf, “CS229 Final Report: Modeling Flight Delays”, Computer Science Department, CS 229, Stanford Univerisy, Standord, CA 2015

[2] Lawson, D. And Cstillo W., “Predicting flight Delays. Technical report”, Computer Science Department, CS 229, Stanford Univerisy, Standord, CA 2012

[3] Bandoyapadhyay, R. and Guerreror, R. “Predicting airline delays. Technical report”, Computer Science Department, CS 229, Stanford university, Stanford, CA, 2012.

[4] Wang, Y., Sridhar, N., Jehlen, R., and Klein, A., “Modeling flight delays and cancellations at the national, regional and aiprot levels in the united states.”, In Eighth USA/Europe Traffic Management Research and Development Seminar (ATM 2009), NASA, Moffet Field, CA, USA and Washington DC, USA and Fairfax, VA, USA, 2009

[5] E.R. Mueller and F.B. Catterji. “Analysis of aircraft arrival and departure delay characterisitics.” In AIAA’s Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical, NASA, Moffet Fieeld, CA, USA, 2002

[6] Y. Tu, M. O. Ball, and W. Jank, “Estimating Flight Departure Delay Distributions - A Statistical Approach With Long-Term Trend and Short-Term Pattern,” *SSRN Electronic Journal*, 2005.

[7] "On-Time Performance", Transtats.bts.gov, 2017. [Online]. Available: https://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=120&Link=0. [Accessed: 07- Dec- 2017].

[8] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," in IEEE Transactions on Information Theory, vol. 14, no. 3, pp. 462-467, May 1968. doi:10.1109/TIT.1968.1054142 keywords: {Approximation methods;Probability functions;Trees}, URL: http://ieeexplore.ieee.org.stanford.idm.oclc.org/stamp/stamp.jsp?tp=&arnumber=1054142&isnumber=22639

[9] "10 Most Weather-Delayed U.S. Major Airports | The Weather Channel", The Weather Channel, 2017. [Online]. Available: https://weather.com/travel/commuter-conditions/news/most-weather-delayed-airports-us-2016. [Accessed: 08- Dec- 2017].

[10] D. training, "Differences between Baum-Welch and Viterbi training", Stats.stackexchange.com, 2017. [Online]. Available: https://stats.stackexchange.com/questions/581/differences-between-baum-welch-and-viterbi-training. [Accessed: 08- Dec- 2017].

Link to GitHub: https://github.com/ZhijieWang/AA228-final