

Avoiding Flight Delays with Supervised Ranking

Dan Garant John Lalor Adam Nelson

Abstract

In this work, we examine the task of selecting flights to minimize the risk of encountering a flight delay. Taking the perspective of a passenger attempting to book a flight one day in advance or one month in advance, we gather and explore a number weather, airport, and air carrier characteristics that can be used to assess the probability of experiencing a delay. For each time horizon under consideration, these features vary in availability and quality, yielding an interesting pair of machine learning tasks. We evaluate a number of supervised learning models for each task, and present results in the form of rankings, consistent with a strategy that could be employed by a flight-booking service.

1 Introduction

In recent years, the U.S. Bureau of Transportation Statistics (BTS) has enabled easy access to a wealth of information on on-time flight performance of domestic air carriers¹. Given the availability of flight characteristics, the task of determining whether a flight will take off and arrive on time is well-suited for analysis with machine learning. Air transportation has become a central aspect of the U.S. culture for both business and recreation, with the number of domestic flights offerings growing steadily for decades. Accompanying the growth in air travel, numerous online services including Google Flights and Expedia provide passengers the ability to select from a variety of available air carriers, flight dates, and origins, when booking a flight. However, these services provide only high-level historic information about carrier-specific patterns in delays, if at all.

Combining historical characteristics of air carriers and airports with weather forecasts and weather trends (or *normals*) would enable flight booking sites to provide more reliable and targeted information about delay risk. Our objective is to use supervised learning models to identify the flights with a low probability of delay. Our analysis will look at flight delay patterns according to two distinct problems: predicting a flight delay for flights leaving tomorrow, and predicting flight delays for flights leaving in one month (e.g. in thirty days). In section 2, we give an overview of the BTS dataset, previous studies or airline delays using this dataset, and supplemental weather data sources we employ. In section 3,

¹<http://www.transtats.bts.gov/>

we examine patterns in delays that can be exploited to construct a probabilistic model of delays. Section 4 formally details our method of evaluation, and section 5 explores the results obtained by many commonly-used supervised learning models of various levels of sophistication.

2 Data Sources

2.1 Bureau of Transportation Dataset

The original idea for the project came from the 2009 American Statistical Association Joint Statistical Meetings Data Expo [cite]. The challenge was to create a graphical summary of the data, which consisted of flight statistics from October 1987 to April 2008. The submissions for this explored a variety of aspects of the data. (Need to elaborate here).

The data for the expo came from the United States Department of Transportation (DOT). The DOT Bureau of Transportation Statistics (BTS) track flight statistics for domestic flights. For the purposes of our project, we focused on flight statistics from the year 2014, as these flights were more recent than the original expo dataset and still provide a large set of data from which we could obtain conclusions. We obtained the data directly from the BTS website.

The BTS data includes a large number of data fields. Below we enumerate the fields that we included in our analysis, along with a description of each:

- Origin Airport and Departure Airport IDs: Unique IDs used to identify the origin and departure airports for each flight in the dataset.
- Departure Date: Date of flight.
- Month, Day of Month, and Day of Week: Additional date fields at varying levels of detail. These were included in an attempt to determine if there were specific days of the week or other specific periods that could be tied to flight delays.
- Carrier ID: Unique ID for the airline of the flight.
- Scheduled Departure and Arrival Time: Scheduled arrival and departure times for each flight, as provided by the Computer Reservation System (CRS)
- Distance: Distance of the flight

2.2 Related Work

As part of the 2009 task, there were a number of submissions that used the dataset in interesting ways. We will briefly discuss the top four submissions for that task here. Wicklin and Allison [2009] used SAS Software to provide a number of visualizations of the data, including departure delay heatmaps by month for each of the years included in the data and historical comparisons of

flight delays, showing that the number of delays varied widely between years. Hofmann *et al.* [2011] found that Newark airport was the worst offender when it came to flight delays, which one author can back up empirically. Wickham [2011] looked at two specific airports, Oakland and San Francisco, and compared number of flights between the two airports, looking at trends in flight numbers in the two airports. Dey *et al.* [2009] modeled the data as a graph in order to predict delays for any origin-destination pair.

2.3 Supplemental Data

In addition to the data provided by the BTS, we wanted to augment the feature set for flights by looking at weather patterns for each of the airports to determine the effect that weather might have on flight delays. For our two tasks, this involved obtaining two new sets of features: detailed weather information for the one-day task, and high level weather normals for the one-month task. The intuition here was that if you are looking to predict flight delays for tomorrow’s flights, a lot of detailed weather information will be available for tomorrow that is most likely very accurate. One month in advance, however, the weather information is less likely to be accurate, so we instead looked at ten-year normals for the flight dates, in order to get an idea of weather one month in advance.

3 Exploratory Analysis

To inform our selection and tuning of supervised learning algorithms for this task, we performed a number of exploratory analyses characterizing the nature of dependence between specific features and the response (departure delay in minutes). First, we investigated characteristics of organizations involved in fulfilling the flight. ...

To gain a high-level understanding of the relationship between the weather features and delays, we fit a generalized additive model [Hastie and Tibshirani, 1990] of delay duration (in minutes) to a linear combination of cubic spline transformations of various weather features. Naturally, there are associations between each of the weather characteristics. For instance, humidity will necessarily be high when it is raining. To isolate the *independent* effects of each of the features on the departure delay, we predicted the conditional expectation of delay D with respect to an individual feature F_i , holding other features at their mean value:

$$E[D|F_1 = \bar{F}_1, F_2 = \bar{F}_2, \dots, F_i = f_i, \dots, F_n = \bar{F}_n]. \quad (1)$$

Then, sweeping across various settings for f_i and various feature F_i yields a sensible explanatory model of delay with respect to weather characteristics. When stratifying by airport and evaluating (1) with respect to wind bearing, we find a strong non-linear relationship. This dependence is likely the result of differing runway orientation at various airports. When wind is blowing *across* a runway, take-off and landing is more challenging, and some pilots may opt to

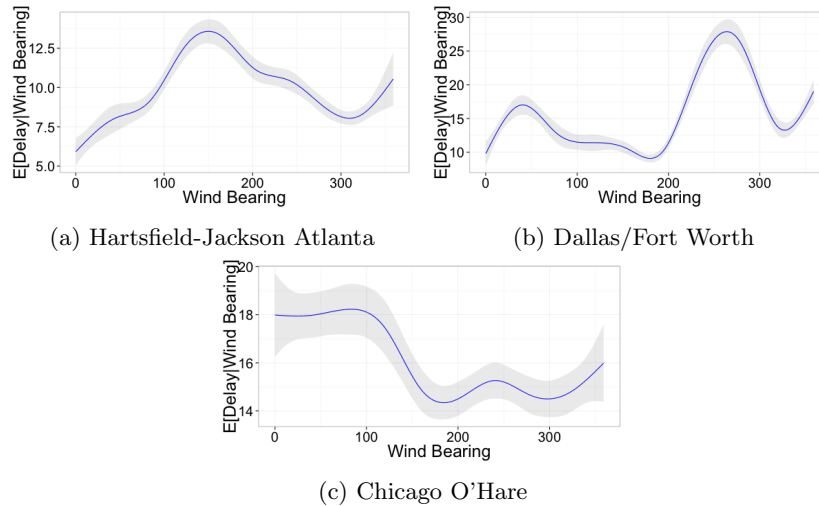


Figure 1: Relationship between wind bearing and expected delays (in minutes) at various airports. One standard error in the conditional expectation is shown in gray.

defer until winds are more favorable. Each airport may have a different runway orientation, so the model must be capable of capturing interactions between a discrete characteristic (airport) and a real-valued characteristic (wind direction).

4 Evaluation Methodology

- Features available in each of the two task “types”
- Discussion of evaluation measures: classification accuracy, log-likelihood, ranking accuracy
- Learning algorithms under consideration
- hyper-parameter optimization

5 Results

References

Tanujit Dey, David Phillips, and Patrick Steele. Minimizing flight delay. Technical report, Tech. rep., Dept. of Math, College of William & Mary, contact authors for manuscript, 2009.

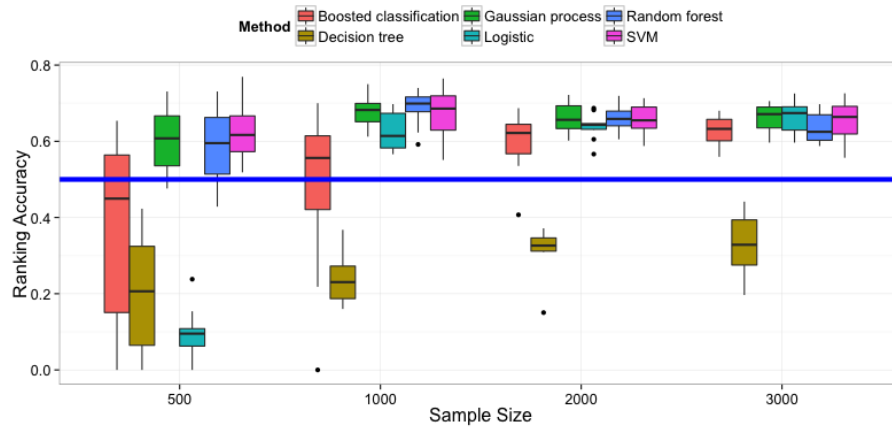


Figure 2: Ranking accuracy of supervised techniques as sample size is varied

Trevor J Hastie and Robert J Tibshirani. *Generalized additive models*, volume 43. CRC Press, 1990.

Heike Hofmann, Dianne Cook, Chris Kielion, Barret Schloerke, Jon Hobbs, Adam Loy, Lawrence Mosley, David Rockoff, Yuanyuan Huang, Danielle Wrolstad, et al. Delayed, canceled, on time, boarding flying in the usa. *Journal of Computational and Graphical Statistics*, 20(2):287–290, 2011.

Charlotte Wickham. A tale of two airports: Exploring flight traffic at sfo and oak. *Journal of Computational and Graphical Statistics*, 20(2):291–293, 2011.

Rick Wicklin and Robert Allison. Congestion in the sky: Visualising domestic airline traffic with sas. *ASA Statistical Computing and Graphics Data Expo*, 2009:14, 2009.