Outline

• **Background**
  – Problem Statement
  – Airline Hub Operations
  – Potential Impact to Flight Delay Reduction

• **Approach**
  – Methodology
  – AAR/ADR Prediction
  – Similar Events

• **Findings**
Background
Major Airlines’ Operations at Hub

• Most airlines concentrate flights at few hubs
  – These hubs account for 40.3% of departures for each airline

• Hubbing Strategies
  – Single Mega Hub with Several Medium Hubs
  – Balanced Hubs
  – Point-to-Point

• It is critical for airlines to streamline operations at airports. Delay at airports can propagate to flights downstream and impact airline significantly.

Airlines Operations at Hub Airports

- Airlines schedule operations based on airport’s VFR capacity rate.
- When airport capacity is reduced from VFR to IFR due to weather (low ceiling and visibility), excessive delay is expected.

The analysis assumes the highest possible VFR rate for ATL. The IFR rate is assumed to be 80% of the VFR based on the highest IFR to VFR ratio from the FAA Airport Arrival Rate (AAR).

Solutions

• Prediction
  – Provides AAR and ADR predictions based on historical weather data, flight demand and airport conditions and limitations for up to 12 hours out

• Similar Events
  – Provides historical AAR and ADR distributions that are similar to the event encountered
Approach
Characteristics of the Problem

- Rare events but critical to operations
  - IMC accounts for 10-20% of the time
- Important to identify both timing and quantity when changes happen
  - When it changes, the quantity it changes, and duration of the change
- Ability to predict future time horizon is critical
- Rates can be dependent by ops at other airports

Change
The AAR of the hour is different from the previous hour
No Change
The AAR of the hour is identical from the previous hour
San Francisco Bay Area East –Flow vs. West Flow
Prediction
Gradient Boosted Machines (GBM)

A family of boosting machine learning techniques using common modeling techniques as weak learners (linear regression, ridge regression, nonlinear splines, decision trees, etc.) which produce a prediction model in the form of an ensemble of weak prediction models.

Unlike aggregation models (e.g. random forests) which take a simple average of models, the main idea of boosting is to add new models to the ensemble sequentially, as weak learners, which are trained with respect to the error of the whole ensemble learnt so far.
Gradient Boosted Machines (GBM)

- To provide a structured and efficient search framework, search parameters need to be supplied to control (tune) the direction of the search, along with the loss function to minimize as the goal.

- Examples of search parameters include early stoppage to avoid overfitting, length of trees, or the number of splines per a predictor, to name a few.

- Data is typically separated into train/test sets where the train set supplies the information to the learning, while the test set is used to score how well the model will generalize into production, given the particular set of search parameters.
Similar Events
Similar Events using KNN

Historic rates selected where historical airport operations and weather conditions are most identical to forecasted airport and weather conditions

• **Most important predictors from xgboost model are considered**

• **Method used for suggestions is K-Nearest Neighbor (KNN)**
  – K is the number of nearest neighbors considered \( \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2} \)
  – Each nearest neighbor is determined by Euclidean distance,
  – Common events posses many close neighbors leading to like suggestions
  – Rare events can lead to distant or fuzzy suggestions
  – Considerations of the value for K adjust for this
  – (more for common events, less for rare events)
Questions?

Ed Cuoco
Ed.Cuoco@weather.com

Alex Huang
Alex.Huang@weather.com