CS5811 Project Proposal:
Flight Delay Prediction based on Bayesian Belief Networks

Jordon Dornbos
Liang Yan
September 18, 2014

1 Project Overview

In this project we are going to develop an iPhone app which could predict a probability that a flight will be delayed or not in a specific future time. Since we do not know how difficult it will be, we first focus on the flight between Chicago and New York, and the prediction time is a day. We will extent the limitation as development continues.

We are motivated to solve this problem because flight delays are a fairly common thing and there seems to be no good source for predicting this. Flight delays are generally only announced about an hour or so before the flight is set to takeoff and they can change unreliably. Our solution can't guarantee 100% accuracy for predicting a delay, but it will attempt to better prepare passengers for any possible delays.

1.1 Problem description

I believe that almost everyone has had a bad experience with a flight being delayed, especially being delayed for something important. According to the U.S. Department of Transportation, over 20 percent of all flights arrive late.[4] It is a major problem for the airplane company and also for us, passengers.

Most of delays are caused by three reasons. First, and most important, is weather. Half of delays are actually caused by weather[2]. Second, it could be some mechanical problem[1], which is mostly decided by the plane features such as the make, model, age and so on. Third, it could be caused by a scheduling problem. If the last flight is delayed, this flight could be delayed also. [3]

1.2 State of the Art

Dr. Rebollo and Balakrishnan[3] proposed amodel uses Random Forest (RF) algorithms, based on temporal and spatial delay, also the local arrival or departure delay situation, they proposed this new network delay prediction model. The predictive performance of the model is evaluated using the 100 most delayed OD pairs in the NAS: the results show that given a 2-hour prediction horizon, the average test error across these 100 OD(origin-destination) pairs is 19% when classifying delays as above or below 60 min.

Dr. Lu and his team predict flight delays by creating a decision tree from a large database. [4]This solution is good at making a simple tree to follow in order to make predictions, but a new tree needs to be created once more data is collected. Using neural networks, however, the graph stays the same, the weights are just learned over time. Our solution will be learning how each input parameter affects the delay probability. This should therefor be more accurate than creating a graph of the most likely outcome based on a series of tests, such as the decision tree provides.

1.3 AI techniques

We plan to use neural network to compute the probability of a delay. The neural network will have multiple inputs describing the the three common issues mentioned above. This includes weather data, information about the plane itself, and the previous flight as well.

We will train the neural network by feeding it data on current flights, and back-propagating the correct value once the flight is over. We will capture data on flights every day and add it to our database. This database will be used to continually train the neural network. Over time it will learn the importance of each input, and should result in very accurate predictions.

Keeping a record of past flight information in our database will also allow us to try out different parameters for the neural network. We can modify parameters, train the network, and keep the parameters that performed best.

2 Tasks

2.1 Platform

iPhone app, Objective-C
Database, MySQL
Model training, Mathlab or write by our own (Python or Objective-c).
Experiment, check the correctness of the prediction.

2.2 Schedule

Preparation:
GUI design, Jordon.
Data collection, Liang

Main: Both
Model construction

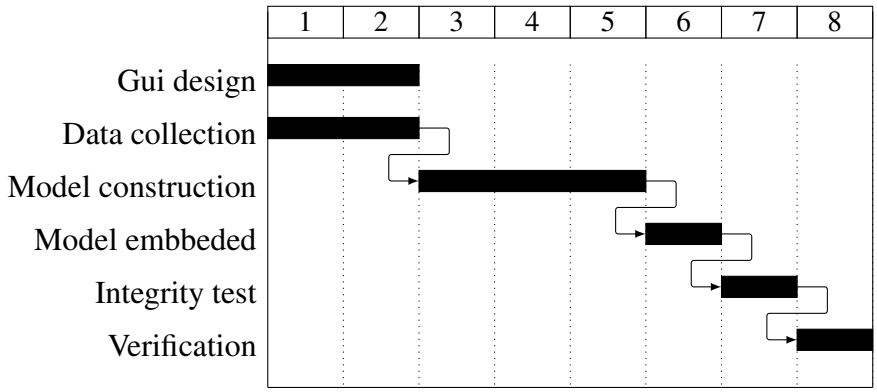
we are still not sure to use MathLab to train model or write our own program to implement it.

Model embedded:
Let the model work well in the iPhone app.

Integrity test:
HCI improvement: Jordon
Function modification: Liang

Verification: Both
Including result analysis and report writing. The experiment is do a prediction based on the history data and check the correctness.

2.3 Arrangement



References

[1] Sina Khanmohammadi, Chun-An Chou, Harold W. Lewis, and Doug Elias. A systems approach for scheduling aircraft landings in jfk airport. In *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, pages 1578–1585, July 2014.

[2] A Klein, C. Craun, and R.S. Lee. Airport delay prediction using weather-impacted traffic index (witi) model. In *Digital Avionics Systems Conference (DASC), 2010 IEEE/AIAA 29th*, pages 2.B.1–1–2.B.1–13, Oct 2010.

[3] Juan Jose Rebollo and Hamsa Balakrishnan. A network-based model for predicting air traffic delays. In *5th International Conference on Research in Air Transportation (ICRAT 2012)*, May 2012.

[4] Lu Zonglei, Wang Jiandong, and Zheng Guansheng. A new method to alarm large scale of flights delay based on machine learning. In *Knowledge Acquisition and Modeling, 2008. KAM '08. International Symposium on*, pages 589–592, Dec 2008.