Flight Delay Prediction Using Neural Networks

Jordon Dornbos, Liang Yan
Michigan Technological University
CS5811 Advanced Artificial Intelligence - Fall 2014

Outline

- What’s the problem
- Our approach
- Data preprocess
- NN model construct
- Data training
- Results
- Future work

What Is the Problem?

What Is the Problem?

1. It happens a lot
2. It costs a lot
3. It is hard to predict

Why Is It Important?

2004: 19 carriers, and 286 airports
total flights: 6987730
delay flights: 2991200 delay rate: 0.428064621844
delay flights: 1727018 delay rate: 0.247150075919

Direct cost of air transportation delay in 2007

Cost Component	Cost (in billions)
Costs to Airlines	$8.3
Costs to Passengers	$16.7
Costs from Lost Demand	$3.9
Total Direct Cost	$28.9
Impact on GDP	$4.0
Total Cost	**$32.9**

The State of the Art

- Airplane delay prediction is a very popular topic, according to the development of Data mining and AI techniques since it has huge history data
- Bayes,SVM,Decision Tree, Random Forest and so on
- We are using neural network

Our Approach

- What kind of data is necessary for our NN??
- The reasons for flight delays?

Our Approach

On-Time Arrival Performance
National (June, 2003 - October, 2014)

on time: 77.86%

air carrier delay: 5.61%

weather delay: 0.74%

national aviation system delay: 6.92%

security delay: 0.05%

aircraft arriving lating delay: 6.83%

cancelled: 1.77%

diverted: 0.22%

Our Approach

According to the reasons, we focused on the weather, NAS , carrier.

1. Carrier:

There are around 30 airplane companies, different company has different type planes, different pilot, different routines, so they could have different delay rate.

2. NAS(national aviation system)

We connect NAS to airports and departure time here. Different airport has different NAS stuff working there, and also different time has airplane working which will let the stuff make different control decision.

Our Approach

3. Aircraft arriving lating delay: 6.83%

   Most direct reason

   Hard to implement.

4. Weather

   we only focus on the weather type and wind speed based on depart airport.

Data Preparation

1. Flight Info **(The Bureau of Transportation Statistics (BTS))**

Year, Month, DayofMonth, DayOfWeek, **DepTime**, CRSDepTime, ArrTime, CRSArrTime, **UniqueCarrier**, FlightNum, TailNum, ActualElapsedTime, CRSElapsedTime, AirTime, **ArrDelay, DepDelay**, Origin, Dest, **Distance**, TaxiIn, TaxiOut, **Cancelled**, CancellationCode, **Diverted**, CarrierDelay, WeatherDelay, NASDelay, SecurityDelay, LateAircraftDelay

2. Weather Info (**QUALITY CONTROLLED LOCAL CLIMATOLOGICAL DATA)**
NOAA's National Climatic Data Center (NCDC)

Wban Number, **YearMonthDay, Time,** Station Type, Maintenance Indicator, Sky Conditions, Visibility, **Weather Type**, Dry Bulb Temp, Dew Point Temp, Wet Bulb Temp, % Relative Humidity, **Wind Speed (kt)**, Wind Direction, Wind Char. Gusts (kt), Val for Wind Char., Station Pressure, Pressure Tendency, Sea Level Pressure, Record Type, Precip. Total

Data Preparation

2004: 19 carriers, and 286 airports
total flights: 6987730
delay flights: 2991200 delay rate: 0.428064621844
delay flights: 1727018 delay rate: 0.247150075919

2005: 20 carriers and 285 airports
total flight: 6992839
delay flights: 3055756 delay rate: 0.436983605657
delay flights: 1808257 delay rate: 0.258586963035

2006: 20 carriers and 289 airports
total flights: 7003803
delay flights: 3204055 delay rates: 0.457473603983
delay flights: 1979441 delay rates: 0.28262374027

Data Preparation

2007: 20 carriers and 304 airports
total flights: 7275289
delay flights: 3459799 delay rates: 0.475554854247
delay flights: 2180386 delay rates: 0.299697510298

2008: 20 carriers and 303 airports
total flights: 6855030
delay flights: 2964145 delay rates: 0.432404380433
delay flights: 1836315 delay rates: 0.267878477556

Data preprocess

1. Carrier

WN,978520,263595,0.269381310551
DL,675724,180546,0.267188970645
NW,500275,109997,0.219873069812
OO,452866,90751,0.200392610618
XE,360569,81762,0.226758262635
CO,296691,67206,0.226518499044
DH,256191,69198,0.270103165217
AS,162825,45496,0.279416551512
B6,89137,21749,0.243995198402
HA,47895,3852,0.0804259317256

AA,684502,171827,0.251024832652
UA,548264,120997,0.220691126902
MQ,464457,127134,0.273726093051
US,412376,98016,0.237685995305
OH,359023,88732,0.247148511377
EV,270844,72735,0.268549423284
HP,192541,54253,0.281773752084
FL,160238,42788,0.267027796153
TZ,74791,16384,0.219063791098

Data preprocess

2. Airport

ATL,817249,253901,0.310677651487	ORD,726724,229395,0.315656287669
DFW,675431,150914,0.223433629786	LAX,459466,99969,0.21757649097
CVG,402517,92586,0.230017614163	IAH,399062,84216,0.211034876786
PHX,336359,90003,0.267580174754	LAS,313019,91861,0.293467808663
EWR,307710,89305,0.290224562088	DEN,307039,62921,0.204928364149
MSP,289384,65726,0.227123821635	DTW,288265,67876,0.235463896068
SLC,273750,59085,0.215835616438	IAD,259031,61782,0.238511992773
BOS,257634,60881,0.236308095981	SFO,255974,59329,0.231777446147
LGA,245772,61397,0.24981283466	PHL,237231,73862,0.311350540191
SEA,215736,57827,0.268045203397	MCO,212738,52734,0.247882371744
CLT,211886,45429,0.214403028043	BWI,209501,50815,0.242552541515
CLE,183278,37621,0.205267407981	DCA,181915,35346,0.194299535497
MDW,181011,48223,0.266409223749	JFK,180046,50615,0.281122602002

Data preprocess

3. Time

0,9690,1932,0.199380804954
1,5573,971,0.174232908667
2,947,188,0.198521647307
3,68,16,0.235294117647
4,27,2,0.0740740740741
5,53570,5500,0.102669404517
6,419905,46095,0.109774830021
7,452430,62075,0.1372035453
8,496550,82451,0.166047729332
9,478199,87748,0.183496828726
10,425542,86003,0.202102260176
11,460790,99912,0.216827622127
12,443487,100702,0.227068662667
13,468127,116688,0.249265690721
14,420975,115277,0.273833363026
15,456742,134950,0.295462208424
16,416803,126922,0.304513163293
17,494172,167962,0.33988570781
18,430814,145739,0.338287520833
19,379793,131068,0.345103780217
20,314814,106579,0.33854593506
21,204765,66958,0.326999243035
22,110348,30245,0.274087432486
23,43598,11035,0.253107940731

Data preprocess

4. Distance

0,1662137,384787,0.231501374435	0-300
1,2162170,540956,0.250191243057	300-600
2,1246760,323163,0.259202252238	600-900
3,1916662,478112,0.24945034648	900~

Data preprocess

5. Weather

weather type

1. SN
2. HZ
3. FG
4. TS
5. RA
6. others

all the data saved in qclcd

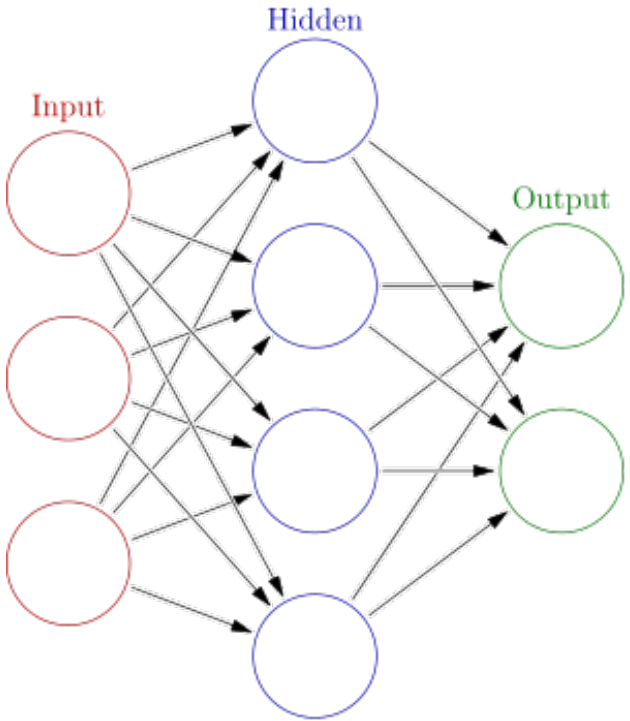
rain, haze, fog, thunderstorm, light snow and snow seem to

have a noticeable impact on flight delays

wind speed:

AI Techniques Used

- Artificial Neural Network
  - Backpropagation of error
  - Normalized prediction data
  - Results in delay probability



Source: Wikipedia

Training and Testing

- Neural network is trained using past flight data
- New values can be entered to get a prediction
- Tested prediction using different past data set

Program Details and Hardware Used

- Written in Python
- Data loaded from CSV files
- Ran on Colossus
  - ~ 15 minutes to train and test

Results Obtained

- Currently: 57.1% accuracy
- Others: ~70% accuracy
- Tweaking neural network parameters and using more data should improve accuracy

Future Work

- Weather data during flight
- Previous flight data
- Rewrite code in a faster language
- Put data in a database

Questions?

Flight Delay Prediction Using Neural Networks

Jordon Dornbos, Liang Yan
Michigan Technological University
CS5811 Advanced Artificial Intelligence - Fall 2014

- What's the problem
- Why is it important
- Existing research
- Challenges
- Our approach
- Results
- Future work