Predicting Delays in Flight Departure Time at SFO

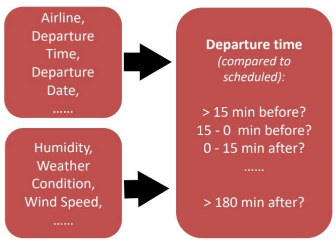
Yutong Sun and Yanqiu Wang, {ycsun, yanqiu}@stanford.edu
CS230 Deep Learning, Stanford University

Abstract

Flight delay is a common issue faced by travelers around the world, and being able to accurately predict flight delays based on external factors will allow travelers to adjust their travel plans ahead of time. In this model, we explored using a **simple NN**, an **LSTM** and a **CNN** model to predict how much a flight would be delayed in 15-minute categorical intervals.

Problem Statement

All of our models in this study address the multi-class classification problem: given information about a flight and the weather data within the hour, which 15-minute departure delay interval is it most likely to fall within?



Data

We synthesized the on-time performance of flights in the United States with the hourly weather data of San Francisco. Sampled data from 2016 was taken, giving us ~10k examples, and each flight is matched to the weather of the hour. Here is an example of one flight entry, each with 13 variables (split for the purpose of this presentation):

DAY_MON	DAY_WEEK	CARRIER_NUM	DEST_STATE	DEP_TIME	DISTANCE
6	5	855	42	1050	2521

Temp	Wind_Speed	Wind_Direction	Humidity	Weather	Pressure	DELAY_GROUP
274	5	20	74	16	1018	-1

Features

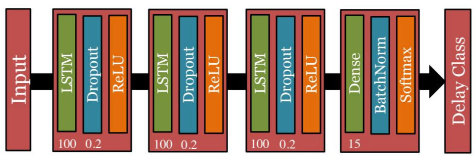
For the simple NN, each training example represents one flight entry with 12 features (all corresponding variables except for delay group). For the LSTM model, each flight's input is the 13 features of its previous flight, including delay group. For the CNN model, each training example is represented by all 13 variables of the 5 previous flights and thus each example's input feature is of dimension (5, 13). All three models use the delay group of the current flight as the ground truth label.

Models

Simple NN

Layer	Parameters	Activation
Hidden Layer 1	W = (25,12) b = (25,1)	ReLU
Hidden Layer 2	W = (12, 25) b = (12, 1)	ReLU
Output Layer	W = (25, 12) b = (25, 1)	Softmax

LSTM



CNN

Layer	f	s	n	p	Activation
CONV	5	1	4	same	ReLU
MaxPool	5	1	/	same	/
CONV	5	1	8	same	ReLU
MaxPool	5	1	/	same	/
CONV	5	2	16	same	ReLU
CONV	5	3	32	same	ReLU
CONV	5	4	64	same	ReLU
FC	/	/	15	/	Softmax

Results

Model	Set	Accuracy (Delay Class)	Accuracy (Binary)	F1 (Binary)
Simple NN	Train m = 88616	45.11%	84.94%	91.80%
	Test m = 6129	36.63%	78.00%	87.50%
LSTM	Train m = 88614	64.09%	89.99%	94.93%
	Test m = 6127	56.70%	83.23%	90.50%
CNN	Train m = 88610	52.71%	89.83%	94.70%
	Test m = 6123	43.75%	81.64%	90.49%

In addition to the accuracy of the delay group prediction, we also calculated the accuracy and F1 score of the model on the binary classification problem of delayed vs. not delayed.

Discussion

Although both LSTM and CNN perform well on the binary task of whether a flight might be delayed, the LSTM model has a much higher accuracy in predicting the actual delay class. We found that during training, the LSTM model overcomes a first plateau at epoch 25 and reaches a lower cost optimum. When stopped at the first plateau, the LSTM has similar performance as the CNN. We believe the CNN model is unable to capture certain information and gets stuck at the first plateau.

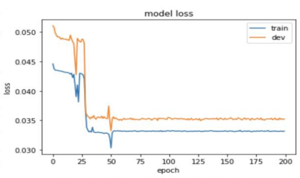


Figure: Model Loss of the LSTM Model During Training

Conclusion & Future

- Both LSTM and CNN model perform well on the binary classification task (>90% accuracy)
- LSTM performs best on multi-class classifying with 57% test accuracy (random guessing only has 6% chance of correctness).
- Future work could look into an LSTM model that further increases the accuracy of multi-class classification.