Challenge #3: Aviation challenge

Context

NAS (National Airspace System) of US is very dynamic and includes a lot of factors that might affect the airline's operations at network level and at the flight level. Those factors include weather, wind, airport capabilities and congestion in different parts of airspace. It is very difficult to plan efficient routes and be sure they will actually be flown.

Your task will be to help airlines predict the time of a flight.

Challenge

In this competition, your goal is to predict the flight time of a flights arriving or departing at O'Hare International Airport in Chicago (ICAO code - **KORD**). You will be given data describing flight plans which were filed pre departure. However flight plans often change both pre-departure and en-route due to changing conditions and restrictions given by Air Traffic Control **(ATC)** in real time. Therefore you know what route is planned, but not what route will actually be flown.

Example situations that might affect the flight time include:

- dangerous weather situation
- unfavourable wind that changes runway configuration at the airport
- congestion in terminal area around destination airport
- actions taken by ATC to delay a plane until the airport is capable of handling the plane

Dataset

Dataset includes:

- output - flight times
- input - flight plans, weather and wind forecasts
- metadata helpful for training.

Each flight has an id assigned to correlate flight plans with their flight times.

You will have:

- **flight_times_labels_train.csv**
  Flight times in seconds you need to predict.

	id	flight_time_s
0	0	3381
1	1	13022
2	2	4846
3	3	4509
4	4	4245

- **flight_plans_train.csv**
  Flight plans you need to predict for the training set. You are given id to correlate this data with **flight_times_labels_train.csv.** Those are flight plans planned pre-departure. Since this is a predictive challenge actual flight plans flown might differ from flight plans planned. Inside this csv you can find:
    - airport pair encoded as for letter ids beginning with "K", e.g. KORD, KLAX
    - fixes, which are points with assigned latitude and longitude used for creating flight plans; take into consideration that some fixes might be nulled, missing etc.
    - decoded fixes, which are the latitudes and longitudes of route fixes
    - assigned altitude in feet, which is the altitude at which the plane is expected to fly
    - requested airspeed in knots, at which an airline is planning to fly

id	departure_airport	arrival_airport	fixes	decoded_fixes	departure_time	assigned_altitude	requested_airspeed
0	KMSP	KORD	KMSP GEP158043 ZZIPR FYTTE KORD	[[44.881972222222224, -93.22177777777777], [44...	2019-09-02 11:47:59 UTC	29000	469.0
1	KORD	KLAX	KORD PEKUE PIPPN ROTTN PWE PWE239044 HIPPI GAB...	[[41.97452222222223, -87.90659722222223], [41....	2019-09-02 13:33:06 UTC	34000	461.0
2	KFAR	KORD	KFAR FAR096018 FAR108025 FAR109025 ZZIPR FYTTE...	[[46.92063888888889, -96.81575], [46.721199612...	2019-09-03 15:11:24 UTC	31000	438.0
3	KORD	KXNA	KORD OBK191036 JOT137013 ARLYN STL SPOKE RZC KXNA	[[41.97452222222223, -87.90659722222223], [41....	2019-09-03 12:57:23 UTC	32000	452.0
4	KORD	KMCI	KORD OBK182032 ACITO ADELL ARLYN COWES343045 C...	[[41.97452222222223, -87.90659722222223], [41....	2019-09-03 12:23:26 UTC	32000	467.0

- **weather_and_winds/**
  Inside you will find npz files in format:
  forecast<forecast-hour>_<time-of-forecast>.npz

  E.g. file forecast4_2019-09-01_01:00:00.npz contains data from forecast done on 2019-09-01 at 01:00:00 for 4h ahead, which is 2019-09-01 at 05:00:00.

Inside you will find Numpy matrices with:

1. **cloud_echo_tops** - those are the levels of clouds in feet; value -999 indicates that there are no clouds
2. **vertically_integrated_liquid_water** - this is an estimate of the total mass of precipitation in the clouds in kg/m^2; you can look at it as the measure of how dangerous it is to fly within the region
3. **flight_level_<altitude-in-feet>_u** - this is wind velocity in m/s at given altitude; last letter (u or v) indicates component of wind vector - u is the zonal velocity (towards east), v is the meridional velocity (towards north)

To understand the geographical location of those matrices you will need lats.npy and lngs.npy.

- **lats.npy & lngs.npy**
  Inside those files you will find latitudes and longitudes of the points in weather and wind matrices. Latitude and longitude matrices have exactly the same shape. All the weather and wind matrices are in Lambert projection.

```
(array([[21.138123  , 21.18688583, 21.23530646, ..., 21.27891197,
         21.23080236, 21.18234976],
        [21.31227557, 21.36117902, 21.40973931, ..., 21.45347062,
         21.40522222, 21.35662986],
        [21.48654735, 21.53559163, 21.58429181, ..., 21.62814912,
         21.5797617 , 21.53102938],
        ...,
        [47.47029682, 47.54181345, 47.61283751, ..., 47.67680663,
         47.60623044, 47.53516035],
        [47.64228127, 47.71395943, 47.78514402, ..., 47.84925781,
         47.77852201, 47.7072913 ],
        [47.81409316, 47.88593295, 47.95727816, ..., 48.0215367 ,
         47.9506412 , 47.87924978]]),
 array([[-122.719528  , -122.53277024, -122.34580087, ...,  -72.82390249,
          -72.63674228,  -72.4497926 ],
        [-122.77209832, -122.58499073, -122.39767026, ...,  -72.77235251,
          -72.58484006,  -72.39753941],
        [-122.8248827 , -122.63742409, -122.44975134, ...,  -72.72059194,
          -72.53272606,  -72.34507324],
        ...,
        [-133.86663443, -133.61178936, -133.35635897, ...,  -61.87573905,
          -61.61977962,  -61.36440376],
        [-133.97309059, -133.71766073, -133.46164092, ...,  -61.77099549,
          -61.51444247,  -61.25847763],
        [-134.08013858, -133.82412204, -133.5675109 , ...,  -61.6656657 ,
          -61.40851711,  -61.15196137]]))
```

- **runway_configs.csv**
  This is a file with runway configuration of all major US airports. You have separate setup of open runways for arrival and departure traffic. Change time indicates when runway change occured.

	airport_icao_code	change_time	arrival_setup	departure_setup
0	KJFK	2019-09-01 01:55:00+00:00	['04R', '04L']	['04L']
1	KJFK	2019-09-02 04:00:00+00:00	['04L']	['04L']
2	KJFK	2019-09-02 07:41:00+00:00	['04R', '04L']	['04L']
3	KJFK	2019-09-02 08:42:00+00:00	['22L', '22R']	['22R']
4	KJFK	2019-09-02 11:48:00+00:00	['22L']	['22R']
5	KJFK	2019-09-02 13:31:00+00:00	['22L', '22R']	['22R']
6	KJFK	2019-09-03 04:55:00+00:00	['04R', '04L']	['04L']
7	KJFK	2019-09-03 18:15:00+00:00	['22L', '22R']	['22R']
8	KJFK	2019-09-05 13:55:00+00:00	['4L']	['04L']
9	KJFK	2019-09-05 19:00:00+00:00	['04R', '04L']	['04L', '31L']
10	KJFK	2019-09-08 01:00:00+00:00	['22L', '22R']	['22R']

Evaluation

Your solutions will be evaluated on a test set. Test set will use the same weather data as train set, but **flight_times_labels_test.csv** and **flight_plans_test.csv** will be published during the event.

Metric used to evaluate your solutions will be mean absolute error of flight time.

CAUTION: Your predictions should be based on weather and wind forecasts not later than departure time. Using forecasts from the future will be considered as cheating. Team which will attempt this will be banned from AiGames for life.

Resources

1. Tutorials about NAS from Federal Aviation Agency (FAA): https://www.faa.gov/about/office_org/headquarters_offices/ang/offices/tc/library/Storyboard/nextgen-overview.html#storyboards
2. Materials about cloud echo tops: http://ww2010.atmos.uiuc.edu/(Gh)/guides/rs/rad/mdr/tops.rxml
3. Quick explanation of *u* and *v* wind components: http://tornado.sfsu.edu/geosciences/classes/m430/Wind/WindDirection.html