A Model Comparison of the US Flight Delay Network (Updated version)

Jiayin Guo, Arianna Groetsema, Quan Chen, Sihui He, Jian Wang, Haochen Li

February 19, 2020

1 Problem Setup

1.1 Data Processing

Asumme there are $n \in \mathbb{Z}^+$ airports from the airport dataset A annually.

1.1.1 Definition of Delayed flights

The scheduled arrival and departure time of flights can be accessed from the dataset.

For clearness, a flight is arrival delayed if the actual arrival time of the flight is greater than the scheduled arrival time 15 minutes. Similarly, a flight is departure delayed if the actual departure time of the flight is greater than the scheduled departure time 15 minutes.

1.1.2 Notation

Let $i \in A$ as an airport, time $t \in T = t_0 + \mathbb{N}dt$. We denote the total number of scheduled arrival flights of airport i at $t \in [t, t + dt)$ as

$$N_{i,t}^{arr} = \{f | \text{ Scheduled Arrival Time of } f \in [t, t + dt)\}.$$

We then denote the total number of scheduled departure flights of airport i at at $t \in [t, t + dt)$ as

$$N_{i,t}^{dep} = \{f | \text{ Scheduled Departure Time of } f \in [t, t + dt)\}.$$

Denote the total number of actual arrival delay flights of airport i at $t \in T$ as

$$D_{i,t}^{arr} = \{f \in N_{i,t}^{arr} | \text{Arrival Delay Time} > 15\}.$$

Denote the total number of actual departure delay flights of airport i at $t \in T$ as

$$D_{i,t}^{dep} = \{f \in N_{i,t}^{dep} | \text{Departure Delay Time} > 15\}.$$

1.2 Definition of Delay Ratio

For an aiport $i \in A$, and time $t \in T$, we let the delay ratio of i at t as

$$p_{i,t} = \frac{d_{i,t}^{arr} + d_{i,t}^{dep}}{n_{i,t}^{arr} + n_{i,t}^{dep}}, \tag{1}$$

where $d_{i,t}^{arr} = |D_{i,t}^{arr}|$,$d_{i,t}^{dep} = |D_{i,t}^{dep}|$, $n_{i,t}^{arr} = |N_{i,t}^{arr}|$,$n_{i,t}^{dep} = |N_{i,t}^{dep}|$

1.3 Goal: prediction of delay ratio

Our goal is to predict the delay ratio of airports by using the past data.

1.4 Description of Training Set and Test Set

1.4.1 Training Set Constrictions

For each time point t in the test set, the information from the time before this time point can be used as the training set for our models.

Fix t(must be chosen from the dataset in 2008) and i, denote t_i^{arr}, t_i^{dep} as the actual arrival time and actual departure time of the flights at airport i particularly. If comparing t with the actual arrival and departure time, each model shall be fitted on a training set under the following constrictions:

(1) If $t_i^{arr}, t_i^{dep} < t$, all the information of both t_i^{arr} and t_i^{dep} can be used.

(2) If $t_i^{dep} < t < t_i^{arr}$, only the information of t_i^{dep} can be used.

(3) If $t < t_i^{dep} < t_i^{arr}$, only scheduled departure and arrival time of the flights at airport i can be used.

1.4.2 Test Set Description

We use the same test sets for all the models. For each test set, we randomly pick 20% of all the airports in 2008, and 20% of all the time points from the original dataset(year 2008). Each time point includes the information of all the airports in the Graphflow object for this specific time.

1.5 Evaluation

MAE(mean absolute error),RMSE(root mean squared error), WAE(weighted absolute error) are used as the evaluation methods to measure the errors produced by our models.

1.5.1 MAE(mean absolute error)

$$MAE = \frac{\sum_i \sum_t |\hat{p}_{i,t} - p_{i,t}|}{\sum_i \sum_t 1} \tag{2}$$

1.5.2 RMSE(root mean square error)

$$RMSE = \sqrt{\frac{\sum_i \sum_t |\hat{p}_{i,t} - p_{i,t}|^2}{\sum_i \sum_t 1}} \tag{3}$$

1.5.3 WAE(weighted absolute error)

$$WAE = \frac{\sum_i \sum_t |\hat{p}_{i,t} - p_{i,t}| n_{i,t}}{\sum_i \sum_t n_{i,t}}, \tag{4}$$

where $n_{i,t} = n_{i,t}^{arr} + n_{i,t}^{dep}$.

1.5.4 WAE(weighted absolute error)

$$WAE = \sqrt{\frac{\sum_i \sum_t |\hat{p}_{i,t} - p_{i,t}|^2 n_{i,t}^2}{\sum_i \sum_t n_{i,t}^2}}, \tag{5}$$

2 Fundamental Model

2.1 Formulation

We train our model by using the dataset in 2007. For an airport i at time t in 2008(the time information includes the date and the hour), we formulate the prediction of the delay ratio of this specific time as response, and the values of the delay ratio of previous $k > 0$ days as predictors. Mathematically, the formulation can be expressed as

$$\hat{p}_{i,t} = \beta_0 + \beta_1 p_{i,t-1} + \beta_2 p_{i,t-2} + ... + \beta_k p_{i,t-k}, \tag{6}$$

where $p_{i,t-k}$ is the delay ratio of the same hour of t in previous k days.

2.2 Data Fitting

We fit our data by multiple linear regression method using the dataset in 2008. However, we randomly pick four airports by their emplacement rankings. We choose LAX(39,636,042/yr), DFW(31,283,579/yr), MEM(2,016,089/yr), AVP(258,830/yr) for two large airports, one middle airport, and one small airport. For these four airports, we apply multiple linear regression on the delay ratio of t and the delay ratio of previous four days of t to observe how significant the value of response would vary if the number of predictors increases.

The fitting results in Figure 1 show that the constant term and the data from previous one day are significant when applying multiple linear regression on the dataset.

When the p value of a predictor is larger than 0.05, the predictor is not well correlated to its response. The coefficients of predictors vary differently for each airport.

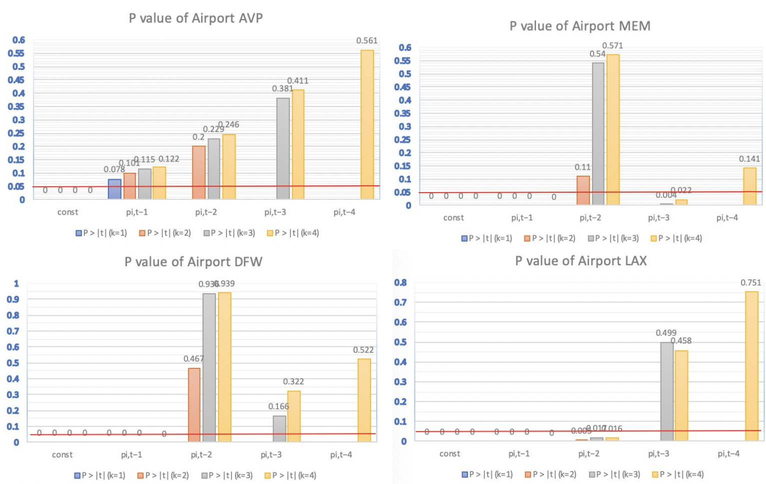

Figure 1: P Values of Multiple Linear Regression for $1 \leq k \leq 4$

k days	MAE	RMSE
k = 1	0.10680970600389689	0.19760764187240015
k = 2	0.1051829267784623	0.1964796070026681
k = 3	0.10407114346343616	0.19586455348387946
k = 4	0.10327180924710248	0.19537780488877401

Table 1: Evaluation Results of the Fundamental Model

In summary, for small airports(enplanements less than 500,000/yr), only constant term β_0 is dominant. For middle airports whose enplanements are around 2,000,000/yr to 5,000,000/yr, the constant term β_0, $p_{i,t-1}$ and $p_{i,t-3}$ are well correlated to the response. For large airports whose enplanements are greater than 20,000,000, the constant term β_0 and $p_{i,t-1}$ are well correlated to the response.

2.3 Evaluation Results

The results are shown in Table 1. Notice that the values of both MAE and RMSE do not differ significantly as we increase k. (STILL NEEDED TO UPDATE WAE)

3 Delay reason based propagation model

3.1 Data Formulation

The following description of Reasons leads to the formulation to this model. For each flight records f there are five reasons accounts for the arrival delay as cited from [https://www.bts.gov/topics/airlines-and-airports/understanding-reporting-causes-flight-delays-and-cancellations]:

(1)Air Carrier: The cause of delay was due to the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).

(2)Extreme Weather(WD): Significant meteorological conditions such as tornado, blizzard or hurricane.

(3)National Aviation System(ND): Delays attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.

(4)Late-arriving aircraft(LD): A previous flight with same aircraft arrived late, causing the present flight to depart late.

(5)Security(SD): Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.

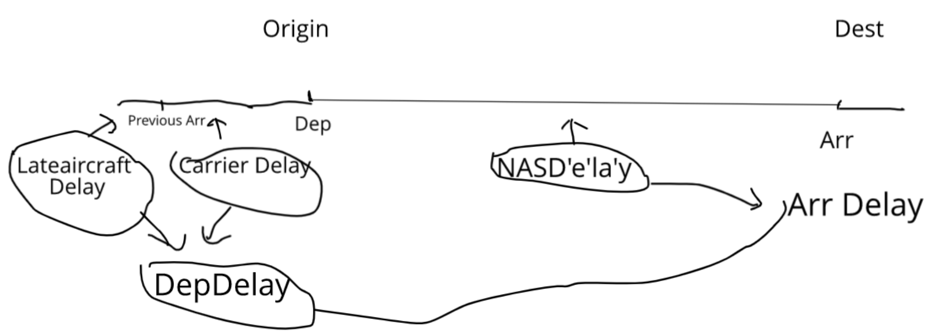

Figure 2: Causal Diagram of Delay Reasons

For simplicity we denote the Delay Reason set by $DR = \{LD, CD, ND, WD, SD, dep, arr\}$ To analysis this more carefully, let us introduce more notations.

3.1.1 History bundle, abstract model ,real model and vectorization

Denote the set of all flight records by $\mathcal{R}$, the set of all aircraft by AC, $\tilde{T} = [t_0, +\infty) \supseteq T$, denote $A \times A \times \tilde{T} \times \tilde{T} \times AC$ by S ,$\mathbb{R}^7$ by D then

$$\mathcal{R} \subseteq S \times D$$

with each $f = (i_o, i_d, t_{dep}^{CRS}, t_{arr}^{CRS}, ac, \delta_{dep}, \delta_{arr}, \delta_{LD}, \delta_{CD}, \delta_{ND}, \delta_{SD}, \delta_{WD}) \in \mathcal{R}$ where i_o is origin airport. i_d is the destination airport, t_{dep}^{CRS} is the scheduled depart time, t_{arr}^{CRS} is the scheduled arrival time,ac is the aircraft for f. δ_{dep} is the departure delay, δ_{arr} is the arrival delay , δ_{LD} is the Late Aircraft delay.

During actually prediction, suppose we want to predict the arr delay of a flight which has not arrived yet at time t, then obviously some data of a flight is not available at time t. To capture this uncertainty. We use the following definition.

Definition 1 *let* $\tilde{\mathcal{R}} = \mathcal{R} \subseteq A \times A \times \tilde{T} \times \tilde{T} \times AC \times (\mathbb{R} \cup \{\infty\})^7 = S \times \tilde{D}, \mathcal{T} \subset \mathbb{R}$
a history map is a map $F : \mathcal{R} \times \mathcal{T} \longrightarrow \tilde{\mathcal{R}}$
with

$$\begin{aligned} F((s, \delta_{dep}, \delta_{arr}, \delta_r), t) &= (s, \delta_{dep}, \delta_{arr}, \delta_r) if t > t_{arr}^{CRS} + \delta_{arr} \\ &= (s, \delta_{dep}, \infty, \infty) if t_{arr}^{CRS} + \delta_{arr} > t > t_{dep}^{CRS} + \delta_{dep} \\ &= (s, \infty, \infty, \infty) if t > t_{arr}^{CRS} + \delta_{arr} \end{aligned}$$

where $s \in S$, $r \in DR$. *And we denote* $F(f, t)$ *by* $F_t(f)$ *and call it the history of* f *at time* t.

Definition 2 *a history bundle* $RF \subseteq \tilde{\mathcal{R}} \times \tilde{T}$ *over record* $\mathcal{R}$ *is a triple* $(RF, R, \pi, \mathcal{T})$ *such that*
(1) $\pi : RF \longrightarrow \mathcal{R}$ *is subjective;*
(2) $\pi^{-1}(f)$ *is bijective to* $\mathcal{T}$

Given a history map F and a record set $\mathcal{R}$, we associate with a canonical record bundle with total space $RF_F = \sqcup_t F_t(\mathcal{R})$, we have
$\pi \circ F_t = id_{\mathcal{R}}$ For canonical history map $F_{\tilde{T}}$, we omit subscript $F_{\tilde{T}}$ in $RF_{F_{\tilde{T}}}$ if it is not ambiguous.

Definition 3 *An attribution* r *is a function over* $S \times D$.

Definition 4 *For an a fixed attribution* r, *an abstract model*

$$\mathcal{M}_r : RF \longrightarrow RVs$$

is a random variable over a history bundle RF.
An abstract model is called normal if it satisfies

$$\mathcal{M}_r(f, t) = r(f) \textit{if } F_t \circ \pi(f, t) = f$$

An abstract model is called static if it factors through $\mathcal{R}$ *and thus factor through* $S_{\mathcal{R}}$, *the static part of* $\mathcal{R}$, *otherwise it is called dynamic.*

Definition 5 *An real model*

$$M : \mathbb{R}^k \longrightarrow RVs$$

is an assignment of a vector x *to* $M(x)$, *where* $M(x)$ *belongs to a fixed family of random variable with parameter depends on* x.

Example 1 *Let ols be the predictor of a fitted Linear Regression model, fixed number* σ , *then* $M(x) \sim N(ols(x), \sigma^2)$ *is a real model.*

Definition 6 *A signature of airports with k fold*

$$\kappa : A^k \longrightarrow \mathbb{R}^{k_1}$$

Definition 7 *A signature of time of k fold*

$$\iota : \tilde{T}^k \longrightarrow \mathbb{R}^{k_2}$$

Example 2 *There is natural or naive signature of airport of fold 1.*

$$\kappa_{na} : A \longrightarrow \mathbb{R}^{|A|}$$
$$a \mapsto e_a$$

Definition 8 *A mapping $v : RF_{F_\mathcal{T}} \longrightarrow \mathbb{R}^n$ is called a vectorization of history bundle $RF_{F_\mathcal{T}}$*

It is called static if it factors through $\mathcal{R}$;

It is called before time $t \in \tilde{T}$, if $\mathcal{T} < t$.

3.1.2 A static model: an example

Consider two normal abstract model $\mathcal{M}_{is\_arrd}$ and $\mathcal{M}_{is\_depd}$, with

$$is\_arrd(f) = \mathbb{1}_{\{\delta arr > 15\}}$$

$$is\_depd(f) = \mathbb{1}_{\{\delta dep > 15\}}$$

We are going to construct their estimations $\widehat{\mathcal{M}}_{is\_arrd}$ and $\widehat{\mathcal{M}}_{is\_depd}$ and at the end of this subsub section we can see they are static and normal.

There are six type signatures considered in the static model: a time-based signatures of fold 1 ι^1,a time-based signatures of fold 2 ι^2,a airport-based signatures of fold 1 κ_{na}, a airport-based signatures of fold 2 κ^2,one delay reason cheating signature $\gamma = id_S : S \longmapsto S$, one delay status based signature (is_arrd, is_depd).

In ι^1, we have six numerical attributes(from pandas.DatetimeIndex): month, day, hour, minute, dayofweek, quarter, and four boolean attributes:

is_month_start, is_month_end, is_quarter_start, is_quarter_end.

Shihui: add iota$_2 by Jiayin$

In κ^{na},we have one numerical attributes: airport index

Remark 1 *Network embedding technique allows us to choose other possible airport-based signatures, we will discuss it further.*

In κ^2, there are two numerical attributes: weight, and distance.

We use popular machine learning classifier to construct our real models

The training set $S^{train}_{vec(sigs),is\_arrd}$ for is_arrd be

$$\{(vec(f), is\_arrd \circ \pi(f)) | (f,t) \in RF_{F_{<20080101}}\}$$

where $vec(sigs)$ is the product $sigs$ of elements in a subset of the following signatures $Sigs$

$$Sigs = \{\iota^1 \circ t_{dep}^{CRS}, \iota^1 \circ t_{dep}^{CRS}, \kappa^{na} \circ i_o, \kappa^{na} \circ i_d, \iota^2 \circ (t_{dep}^{CRS}, t_{arr}^{CRS}), \kappa^2 \circ (i_o, i_d)\} \circ \pi$$

Let M_{ML}^{sigs} be the real model, which is the predictor of machine learning classifyer ML trained by $S_{vec(sigs),is\_arrd}^{train}$.

For all subsets in $Sigs$ and all machine learning classifiers$CLS = \{NN, DT, LR, RF\}$ including neural network, decision tree, logistic regression, random forest. We build $2^6 * 4$ static models.

For $\mathcal{R}_1 = \{f \in \mathcal{R} | o_i = LAX, o_d = MEM\}$, $\kappa^{na} \circ i_d, \iota^2 \circ (t_{dep}^{CRS}, t_{arr}^{CRS}), \kappa^2 \circ (i_o, i_d)$ are constants, so it reduced to $2^3 * 4$ models. We set $\widehat{\mathcal{M}}_{is\_arrd}$ to be the predictor of the best model in the model set $Sig^{\mathcal{P}} \times CLS$ with $2^3 * 4$ elements.

Similarly, we set $\widehat{\mathcal{M}}_{is\_depd}$ be the be the predictor of the best model in the model set $Sig^{\mathcal{P}} \times CLS$ trained by $S_{vec(sigs),is\_depd}^{train}$

In detail we split $S_{vec(sigs)}^{train}$ as training set and validation set; perform a 5 fold cross validation; and measure by *accuracy* metric for each validation set.

We find the tops models are ??, with accuracy ?? on the validation set.

By training the models with $S_{vec(\gamma)}^{train}$,where γ is the cheat signature, we showed that all estimated model are normal. They are static by construction.

3.1.3 Dynamic model

The reason to consider abstract models over RF instead of over $S_{\mathcal{R}}$ is because of the no treating principle.

and the observation of the following fact

$$corr(\mathbb{1}_{\{\delta_{arr} > 15\}}, \mathbb{1}_{\{\delta_{dep} > 15\}}) = 0.83$$

The following scenario is a motivating one.

for a flight f with scheduled arrival time at $t_{CRS}^{arr} = 2018/3/212 : 30 : 00$ and scheduled departure time at $t_{CRS}^{dep} = 2018/3/211 : 30 : 00$ from LAX to ATL , we know $f \in N_{i,t}^{arr}$ with $i = ATL$, $t = 2018/3/212 : 00 : 00$. If we know by the time t it is departed delayed ,then the best guess for $\mathcal{M}_{is\_arrd}(f, t)$ under rwse metric will be a constant $E(\mathcal{M}_{is\_arrd}(f, t))$ so far the best gust is something related to 0.83, which will be much higher than the average between $12 : 00 : 00 - 13 : 00 : 00$ from LAX to ATL which is 0.32. as noted is_arrd relies a dynamic attribution.

The second point we need to consider dynamic is that suppose now δ_{dep} is not available at t then it implies that $\delta_{dep} > 30$, which almost gurantee a arrival delay for f

Our model will based on this principle: as the predict time increase, the uncertainty will decrease. Maybe I shall call it negative entropy model... It has also something to do with martingale theory

3.1.4 Reason based dynamic model

If we consider d_{it}^{X} as a times series TS_G over a network ,that is for each node i and each $t \in T$, we have assigned a number d_{it}^{X} . The nature of our problem is actually a generalized arrival process GA_G over a network, that is for each particle 'arrival' each nodes, it has a numeric feature as dep delay or arrival delay

From a generalized arrival process over a network to a times series over a network. we do two operations.

$$GA_G \longrightarrow A_G \longrightarrow TS_G$$

The fundamental model and RNN works on TS_G We can not expect too much for the preciseness of this type of prediction, as two operation has applied to the original GA_G.

The dynamic model and static model works on A_G, as only one operation has applied to GA_G to get A_G. we are working on data with less error introduced than The fundamental model and RNN does.

The reason based dynamic model will based one the following fact and working on GA_G

$$\mathcal{M}_{depD}(f,t) = \mathcal{M}_{LD}(f,t) + \mathcal{M}_{CD}(f,t) + \mathcal{M}_{WD}(f,t) + \mathcal{M}_{SD}(f,t)$$
$$\mathcal{M}_{arrD}(f,t) = \mathcal{M}_{depD}(f,t) + \mathcal{M}_{ND}(f,t) + \mathcal{M}_{WD}(f,t) + \mathcal{M}_{SD}(f,t)$$

Based on the following correlation matrix of all data availe in 2007, SAY

	WeatherDelay	SecurityDelay	NASDelay	LateAircraftDelay	CarrierDelay
WeatherDelay	1.000000	-0.005225	0.000416	-0.043826	-0.056950
SecurityDelay	-0.005225	1.000000	-0.016929	-0.015118	-0.014288
NASDelay	0.000416	-0.016929	1.000000	-0.144510	-0.140699
LateAircraftDelay	-0.043826	-0.015118	-0.144510	1.000000	-0.105054
CarrierDelay	-0.056950	-0.014288	-0.140699	-0.105054	1.000000

Figure 3: Delay Reasons

MORE COMMENTED BY JIAYIN we introduce our assumption.

Assumption 1 *For a fixed $f \in \mathcal{R}$, if δ_{arr} and δ_{dep} are both unavailable, then ($\mathcal{M}_{LD}(f)$,$\mathcal{M}_{CD}(f)$,$\mathcal{M}_{ND}(f)$,$\mathcal{M}_{WD}(f)$,$\mathcal{M}_{SD}(f)$) are independent.*

Difficulty for the reason based dynamic model
base on $rwse$ metric the best guess for $\mathcal{M}_{is\_arrd}$ $E(\mathcal{M}_{is\_arrd}(f,t))$
So in the context of this model we need to estimate

$$E(\mathbb{1}_{\{\mathcal{M}_{depD}(f,t)+\mathcal{M}_{ND}(f,t)+\mathcal{M}_{WD}(f,t)+\mathcal{M}_{SD}(f,t)>15\}})$$

So it is not enough to construct our estimated abstract model e.g. $\widehat{\mathcal{M}}_{ND}(f,t)$ as a number valued, that is $E(\mathcal{M}_{ND}(f,t))$.As

$$E(\mathbb{1}_{\{\mathcal{M}_{depD}(f,t)+\mathcal{M}_{ND}(f,t)+\mathcal{M}_{WD}(f,t)+\mathcal{M}_{SD}(f,t)>15\}})$$
$$\neq E(\mathbb{1}_{\{E(\mathcal{M}_{depD}(f,t))+E(\mathcal{M}_{ND}(f,t))+E(\mathcal{M}_{WD}(f,t))+E(\mathcal{M}_{SD}(f,t))>15\}}$$

$\widehat{\mathcal{M}}_{ND}(f,t)$,$\widehat{\mathcal{M}}_{LD}(f,t)$,$\widehat{\mathcal{M}}_{CD}(f,t)$ actually need to be distribution valued, this add to the estimation error.

below is not ready

below is not ready

below is not ready

below is not ready

below is not ready

below is not ready

below is not ready

below is not ready

For example we will construct some estimated good abstract model $\mathcal{M}_r(f,t)$ at the end of this section. For notation simplicity, we omit f,t if it is not ambiguous.

Now we are ready to formulate our first layer of our mixed model. Based our interpretation for the delay reasons. We assume a family of good abstract models satisfies

$$\mathcal{M}_{depD}(f,t) = \mathcal{M}_{LD}(f,t) + \mathcal{M}_{CD}(f,t) + \mathcal{M}_{WD}(f,t) + \mathcal{M}_{SD}(f,t)$$
$$\mathcal{M}_{arrD}(f,t) = \mathcal{M}_{depD}(f,t) + \mathcal{M}_{ND}(f,t) + \mathcal{M}_{WD}(f,t) + \mathcal{M}_{SD}(f,t)$$

where $pre : \mathcal{R} \longrightarrow \mathcal{R}$ maps a flight f to a flight $pre(f)$ which most recent flight that aircraft of f is being used.

$$d_{it}^{\chi} = \sum_{f \in N_{it}^{\chi}} \mathbb{1}_{\{\mathcal{M}_{\chi}(f,t)>15\}}$$

with $\chi \in \{arrD, depD\}$

Suppose we construct some estimated abstract model

$$\widehat{\mathcal{M}_{arrD}}, \widehat{\mathcal{M}_{depD}}, \widehat{\mathcal{M}_{LD}}, \widehat{\mathcal{M}_{CD}}, \widehat{\mathcal{M}_{ND}}, \widehat{\mathcal{M}_{WD}}, \widehat{\mathcal{M}_{SD}}, \widehat{\mathcal{M}_{LD}}$$

then we some simplifications.

$$\mathcal{M}\widehat{_{WD}(f},t) = 0$$
$$\mathcal{M}\widehat{_{SD}(f},t) = 0$$

Though this obvious make $\widehat{\mathcal{M}_{WD}}$ and $\mathcal{M}\widehat{_{SD}(f},t)$ biased. But based on following chart, WD and SD does not contribute very much to our prediction. So it is a trade over between biased and variance.

WeatherDelay	0.073896
SecurityDelay	0.002186
NASDelay	0.533450
LateAircraftDelay	0.457368
CarrierDelay	0.418452

And our predict for

$$\widehat{d_{it}^{\chi}} = \sum_{f \in N_{it}^{\chi}} \mathbb{1}_{\{\widehat{\mathcal{M}_{\chi}}(f,t) > 15\}}$$

with $\chi \in \{arrD, depD\}$ gives the prediction

$$\hat{p}_{i,t} = \frac{\widehat{d}_{i,t}^{arr} + \widehat{d}_{i,t}^{dep}}{n_{i,t}^{arr} + n_{i,t}^{dep}}.$$

Remark 2 *I need say something about why I choose model to be random variable valued. by Jiayin*

It remains to construct $\widehat{\mathcal{M}_{CD}}, \widehat{\mathcal{M}_{ND}}, \widehat{\mathcal{M}_{LD}}$ to finish our construction for the whole hyper model.

Before we get into a more detailed description of this estimated abstract models. Let us introduce more notation.

Suppose we have some real models $M_{CD}, M_{ND}, M_{LD}, M'_{LD}$ described as below

For $f = (i_o, i_d, t_{dep}^{CRS}, t_{arr}^{CRS}, ac, \delta_{dep}, \delta_{arr}, \delta_{LD}, \delta_{CD}, \delta_{ND}, \delta_{SD}, \delta_{WD}) \in \mathcal{R}$,

M_{LD} has input δ_{arr} of $pre(f)$, $\kappa(i_o), \iota(t_{dep}^{CRS})$ and output estimated LD delay distribution $\delta_{\hat{LD}}$ It is trained by data satisfying

$$(\delta_{arr} + t_{arr}^{CRS})(pre(f)) < t_{dep}^{CRS}$$

M'_{LD} has input δ_{arr} of $pre(f)$, $\kappa(i_o), \iota(t_{dep}^{CRS})$ and output estimated LD delay distribution $\delta_{\hat{LD}}$ It is trained by data satisfying

$$(\delta_{arr} + t_{arr}^{CRS})(pre(f)) > t_{dep}^{CRS}$$

M_{CD} has input $\kappa(i_o), \iota(t_{dep}^{CRS})$ and output estimated CD delay distribution $\delta_{\hat{CD}}$

M_{ND} has input $\kappa(i_d), \iota(t_{dep}^{CRS}, t_{arr}^{CRS})$ and output the estimated ND delay

Now we are ready to construct $\widehat{\mathcal{M}_{CD}}, \widehat{\mathcal{M}_{ND}}, \widehat{\mathcal{M}_{LD}}$

$$
\begin{aligned}
\widehat{\mathcal{M}_{LD}}(f,t) =& M_{LD}(\delta_{arr}(pre(f)), \kappa(i_o), \iota(t)) \text{ if } F_t(pre(f)) \neq \infty \\
\widehat{\mathcal{M}_{LD}}(f,t) =& M'_{LD}(\delta_{arr}(pre(f)), \kappa(i_o), \iota(t)) \text{ if } F_t(pre(f)) = \infty \\
\widehat{\mathcal{M}_{CD}}(f,t) =& M_{CD}(\kappa(i_o), \iota(\widehat{\mathcal{M}_{LD}}(f,t) + t)) \\
\widehat{\mathcal{M}_{ND}}(f,t) =& M_{ND}(\kappa(i_d), \iota(t_{dep}^{CRS} + \delta_{dep}, t_{arr}^{CRS} + \delta_{dep})) \text{ if } F_t(\delta_{dep}) \neq \infty \\
\widehat{\mathcal{M}_{ND}}(f,t) =& M_{ND}(\kappa(i_d), \iota(t_{dep}^{CRS} + \widehat{\mathcal{M}_{LD}}(f,t) + \widehat{\mathcal{M}_{CD}}(f,t), \\
& t_{arr}^{CRS} + \widehat{\mathcal{M}_{LD}}(f,t) + \widehat{\mathcal{M}_{CD}}(f,t))) \text{ if } F_t(\delta_{dep}) = \infty
\end{aligned}
$$

3.2 Data Fitting

We fit our data by multiple linear regression method based on

3.3 Evaluation Result

4 RNN Model