

Chained Predictions of Flight Delay Using Machine Learning

Jun Chen^{*},

San Diego State University, San Diego, California, 92182

Meng Li[†]

Purdue University, West Lafayette, Indiana, 47906

Flight delay creates major problems in the current aviation system. Methods are needed to analyze the manner how delay propagates in the airport networks. Traditional methods are inadequate to the task. This paper presented a new machine learning based air traffic delay prediction model that combined multi-label random forest classification and approximated delay propagation model. To improve the prediction performance, an optimal feature selection process is introduced and demonstrated to have better performance than directly using all the features of available datasets. Departure delay and late arriving aircraft delay are shown to be the most important features for delay prediction. To utilize these two features, a delay propagation model is proposed as a link to connect them to build a chained delay prediction model. Given the initial departure delay, the chained model is demonstrated to have the ability to predict the flight delay along the same aircraft's itinerary. By updating the actual departure delay with the iteration number along the itinerary, the model's accuracy can be further improved. Our application results clearly demonstrate the value of machine learning and delay propagation for analyzing and predicting the air traffic delay in daily operation.

I. Introduction

With rapid growth of air traffic, increasing flight delays in the United States (US) have become a serious and prominent problem. According to the Bureau of Transportation Statistics (BTS), nearly one in four airline flights arrived at its destination over 15 minutes late [1,2]. It is reported that the annual total cost of air transportation delays was over \$30 billion, which poses a significant challenge to the development of Next Generation Air Transportation System (NextGen) [1,3,4]. This fact motivates the need for accurate and practical prediction of flight delays, especially for individual flights.

Beyond the delays directly created by limited airspace capacity, a third of the late arrivals were caused by an aircraft arriving late and thus having to depart late on its next flight, which is known as delay propagation [1,5]. Given the fact that airlines fly their aircraft on daily scheduled itineraries that require visits to a sequence of airports, the late-arriving aircraft delay early in the day has a significant impact on the downstream delay performance [6]. For example, if an aircraft is delayed by one hour in departure from the first airport, it will almost certainly be late in arriving at its next airport; the late arrival may also result in a late subsequent departure of that aircraft, which will lead to a sequence of late-arriving aircraft delays [5]. The delay propagation is inherent with the National Airspace System (NAS), which includes a large number of connective resources, such as aircraft, crew, passengers and gate space. Moreover, the increasing air traffic demand pushes the NextGen to reduce slack time between arrivals and departures, which will make the NAS further suffering from serious delay propagation through the network. Therefore, the modeling of delay propagation is a key factor for the success of accurate flight delay prediction.

Delay analysis has been an important research topic, which attracts a great deal of attention to delay prediction. Mueller et al. [7] demonstrated the characterization and distribution of the delay in a traditional statistical approach. Klein et al. [8] integrated the convective weather forecasts, terminal airports weather forecasts and scheduled flights information to predict the daily airport delay time based on a metric called Weather Impacted Traffic Index (WITI). Considering delay propagation, Pyrgiotis et al. [5] proposed an Approximate Network Delays model (AND), which analyses the delay propagation within the network of airports. By providing the demand and capacity of the airports and flight itineraries, AND computes a queuing model and a delay propagation model to analyze the delay propagation phenomenon within 30 main airports of US. These models are able to predict the flight delays within the airport systems, but they can only compute the macro daily or hourly delay and fail to predict the individual flight delay.

Apart from analyzing the delay from building a mathematical model, machine learning has also drawn a lot of attention. By training the linear model in the Bayes network with cancellations, the Bayesian network model can predict the delay propagation within three major airports [9]. Based on random forest classification and regression,

^{*}Assistant Professor, Aerospace Engineering Department, AIAA Member jun.chen@sdsu.edu

[†]Graduate Student, School of Aeronautics and Astronautics, li2607@purdue.edu

Rebello et al. [10] captured the macro delay patterns and dependencies of airports within the network of air traffic systems. However, to predict the departure delay, they considered the taxi in/out and wheel on/off time. These features are inaccessible in actual practice, especially when predicting future individual flights. For predicting arrival delays of individual flights, Choi et al. [11] proposed a machine learning based model focused on weather-induced delay prediction of individual flight. By combining the weather data and air traffic data, this model improves the binary classification accuracy for specific individual origin-destination pair. However, this model is limited to provide only binary classification of delay. Moreover, most of the previous machine learning work have a problem of feature selection.

Overall, most of the previous work has limitations. The traditional approaches can model the delay propagation of air traffic operation, but they lack the ability to analytically analyze the huge volume of traffic and weather dataset. On the other side, the machine learning approaches has the ability to discover the hidden patterns in the data, but they lack the inner relationships between different airports, such as delay propagation. Therefore, a mixed approach including the machine learning and air traffic operation is proposed with feature selection to improve the performance of the prediction. Moreover, this new approach can work as a chain model to provide sequence future delay predictions for individual flights along their scheduled itineraries.

The major contributions of this paper can be summarized in three parts. First, a feature selection process is introduced for the multi-label classification algorithm, which supports both the departure and arrival delay prediction modules. The predictable factors that can affect the air traffic include the weather data and air traffic performance data. Based on the combinations of Bureau of Transportation Statistics (BTS), the National Oceanic and Atmospheric Administration (NOAA) and Aviation System Performance Metrics (ASPM), we design an algorithm to select the optimal training features that predicts the departure delay and arrival delay with the highest accuracy. Second, a delay propagation model is proposed with a key concept as Late Arriving Aircraft Delay (LAAD), which is one of the five main causes of airline delays in BTS. The LAAD describes the inner relationship between a previous arrival delay and the present departure delay using the same aircraft. Finally, a chain model is built by using the LAAD as the link connecting previous arrival delay and present departure delay. Given the initial departure delay, the chain model works by iteratively predicting arrival delay, computing LAAD, and predicting next departure delay along the same aircraft's itinerary. By updating the actual departure delay with the iteration number, the model's accuracy can be further improved.

The rest of this paper is organized as follows. Section II introduces a general description of the chain model, outlines its overall frame and presents the details of delay propagation model. Section III describes how the raw data is processed. Section IV explains the details of feature selection. Section V presents application results to show the performance of the chain model. Section VI concludes this paper.

II. Model Description

The mixed approach for predicting arrival delay and departure delay of individual flights includes three modules: the arrival delay prediction module, the departure delay prediction module and the delay propagation module. The frame of the mixed approach is shown in the Figure 1. The delay prediction model works as the link, which connects the arrival delay prediction module and the departure delay prediction module. The chain model can be built by iteratively running the connected modules when the initial departure delay is given. All the other inputs include the flight schedule, and training set features. The arrival delay and departure delay prediction modules are based on random forest model trained with selected features, where the detail of feature selection is discussed in section IV. The delay propagation module is an optimized function to fit the historical LAAD.

A. Random Forest Classifier

The Random Forest (RF) classifier is an ensemble method based on multiple decision trees [12]. By combining the Bootstrap aggregating [13] and random space method [14], RF overcomes the drawbacks of individual decision tree. RF is widely used in industry because it can classify high dimensional data in short time with good performance and it has low sensitivity to outliers in the training data [13].

Moreover, RF was chosen as the core for our prediction modules for two reasons. First, RF is tested to have superior performance than other classification models [10, 11]. Second, RF can output the importance of the features in its learning process. which is the key for our feature selection process in section IV.

B. Delay Propagation Model

Some key parameters and variables are defined to help describe the delay propagation model as follows:

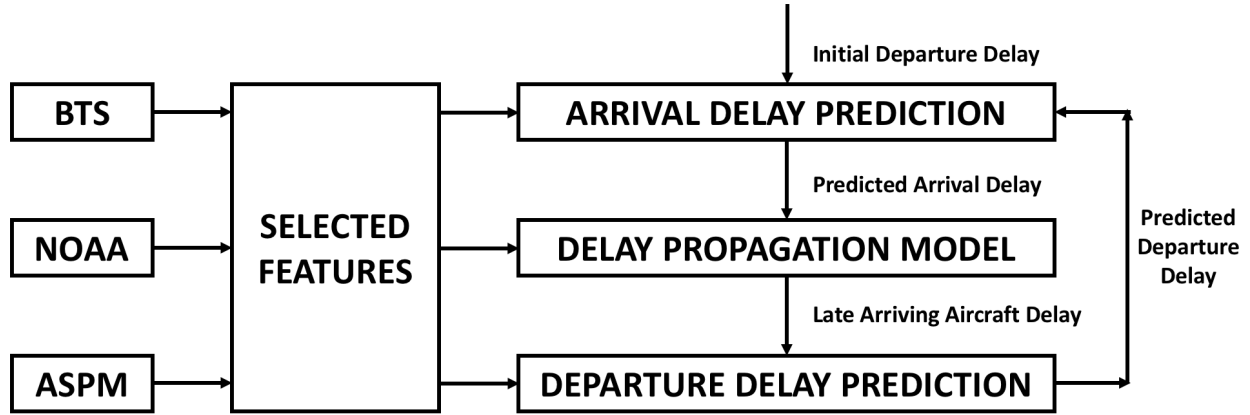


Figure 1. Mixed approach for chained delay prediction.

- f = a flight in an aircraft's itinerary
- f' = the immediate predecessor flight of f
- t_{min} = minimum airport turnaround time in minutes, which is unknown from the data
- $t_s(f, f')$ = scheduled airport turnaround time between f and f' in minutes
- $t_{LAAD}(f)$ = late arriving aircraft delay of flight f in minutes
- $t_a(f)$ = arrival delay of f in minutes
- $SD(f)$ = scheduled departure time of f
- $SA(f)$ = scheduled arrival time of f

Delay propagation in the air traffic system can be divided into two parts, propagation through the en route link between the airports and the ground propagation inside the airport. Although the previous RF classifier can predict the departure delay and arrival delay separately, the transition from the actual arrival delay to the LAAD is not clear. According to BTS, the LAAD is one of the five main causes of airline delays and it describes the inner relationship between the previous arrival delay and the present departure delay using the same aircraft. To compute the LAAD, it is assumed that LAAD $t_{LAAD}(f)$ is a function of the actual arrival delay $t_a(f)$ and scheduled airport turnaround time $t_s(f, f')$. The formula is given as the following:

$$t_{LAAD}(f) = t_a(f) - (t_s(f, f') - t_{min}) \quad (1)$$

$$t_s(f, f') = SD(f') - SA(f) \quad (2)$$

where scheduled airport turnaround time $t_s(f, f')$ is the time period that the aircraft will stay in the airport. This value is not provided by the BTS and need to be computed from the scheduled departure time of the immediate predecessor flight f' minus the scheduled arrival time of flight f . Minimum airport turnaround time (t_{min}) is the minimum necessary time for the aircraft to stay at the airport. Normally the flight needs to complete the unloading/loading, cleaning, aircraft fueling, safety inspection, etc. We assume the minimum turnaround time is a constant for all the flights in all the airports of our air traffic system. By adjusting the minimum turnaround time, we can fit an optimized approximation function to minimize the error with the actual LAAD.

In fact, we find that there is some negative scheduled airport turnaround time, which means the $SD(f')$ is earlier than the $SA(f)$. These negative values account for a small proportion for all the flight data, but they are responsible for the emergence of the carrier and NAS delay. The following Figure 2 gives the average departure delay distribution with the scheduled airport turnaround time. In Figure 2, horizontal axis is the scheduled airport turnaround time intervals from -60 to 750 minutes. Left vertical axis is the average departure delay time and the right vertical axis is the total number of the flights in the corresponding time intervals. The departure delay distribution follows the pattern with minimized value at 30-90 time intervals with the most flights. Although only a few proportion of flights fall out of the 0-120 time intervals, their averaged departure delay time is high. When the turnaround time is negative, the flight does not have the spare time to absorb the delay and may cause further delay instead. When the turnaround time is large, there is a slight rise of the departure delay. A possible explanation is that the flight stays at the airport too long and has the corresponding delay.

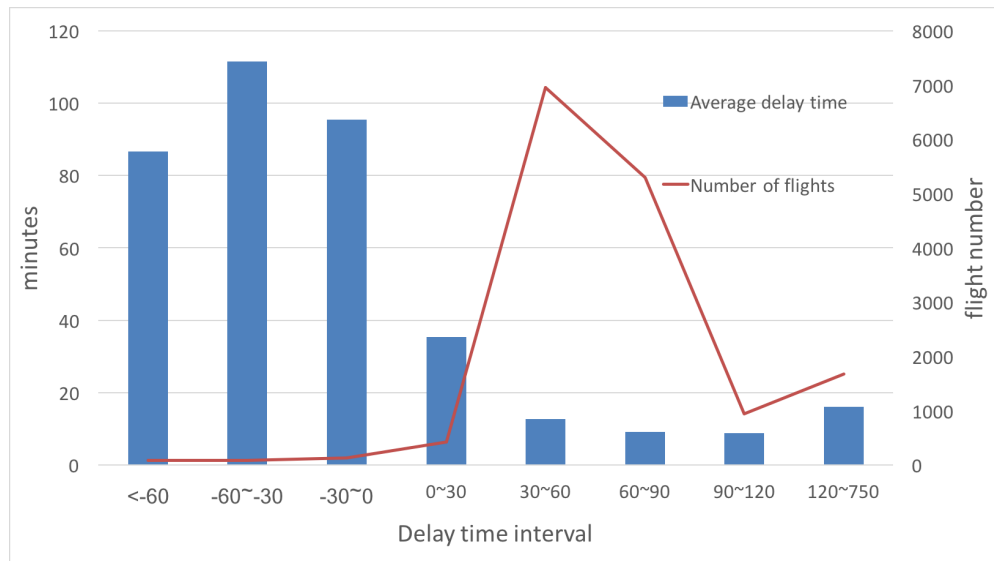


Figure 2. Departure delay distribution with scheduled airport turnaround time.

To fix the negative airport turnaround time issue, a formula to compute the LAAD is added with the MAX function.

$$t_{LAAD}(f) = \max(t_a(f) - \max(t_s(f, f'), 0), 0) + t_{min}, 0) \quad (3)$$

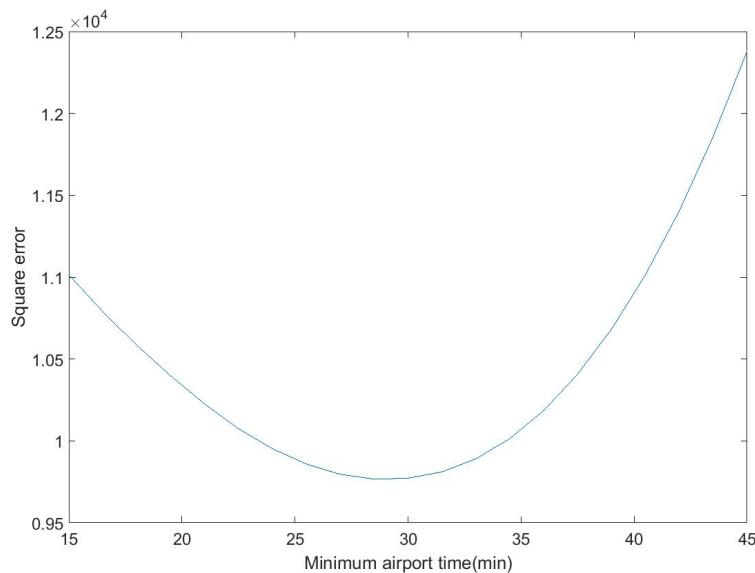


Figure 3. Approximation error with different minimum airport turnaround time.

The LAAD approximation errors based on the propagation formula is shown in Figure 3 as a function of the minimum airport turnaround time. This best fitting problem is actually a simple convex optimization problem. To minimize the squared approximated delay error, the optimal minimum turnaround time is chosen to be 28 minutes. This minimum airport turnaround time is applied in the later model.

III. Data Analysis

A. Source of the data

Three databases are chosen to select the best combination of the features based on our previous data analysis experience [15]. The databases are airline on-time performance database of the Bureau of Transportation Statistics (BTS)^a, Local Climatological Data (LCD)^b at the National Oceanic and Atmospheric Administration (NOAA) and Aviation System Performance Metrics (ASPM)^c.

The BTS provides the air traffic on-time performance data reported by certified US. Air carriers. To explain the arrival delay of the flights, causes of the delay are reported in five categories: air carrier, extreme weather, national aviation system, late arriving aircraft and security. LCD contains weather summaries for major airports that include a daily account of temperature extremes, degree-days, precipitation amounts, winds and special weather. ASPM data contains airport capacity and throughput data for main airports of US. In our model, only the hourly departures and hourly arrivals are applied in corresponding airports.

B. Preprocessing of the data

Air traffic data are selected from the database for the Chicago O'Hare International Airport (ORD) as origin or destination airport from 2016 July to 2017 June. Instead of the precise delay time, we use the delay group in our research for random forest classification. The application of the delay group is suitable for the uncertainty of the actual air traffic control. The delay groups are divided by 15 minutes from -2 to 12. All the non-positive delay groups are classified as on-time flights. The maximum of the delay group is 12 (delay time greater than 180 minutes)^d. The Details of the delay groups are shown in Table 1.

Table 1. Reference table of delay groups

Delay groups	Delayed time (min)
-2	(-inf, -15)
-1	[-15,0)
0	[0,15)
1	[15,30)
2	[30,45)
3	[45,60)
4	[60,75)
5	[75,90)
6	[90,105)
7	[105,120)
8	[120,135)
9	[135,150)
10	[150,165)
11	[165,180)
12	[180,+inf)

The following delay-related data are extracted from the BTS database.

- Day of month
- Day of week
- Scheduled departure time
- Scheduled arrival time
- Scheduled elapsed time
- Departure delay group
- LAAD group: Divided by 15 minutes into groups from -2 to 12

^a<https://www.transtats.bts.gov>

^b<https://www.ncdc.noaa.gov/cdo-web/datatools/lcd>

^c<https://aspm.faa.gov/apm/sys/AnalysisAP.asp>

^dThe definition of delay group can be found on BTS (<https://www.transtats.bts.gov/Fields.asp>)

- Arrival delay group

The following delay-related weather data are extracted for the origin and destination airports from the NOAA database.

- Hourly Visibility
- Hourly Present Weather Type: Special weather type including drizzle, mist, thunderstorm/snowstorm and other special weathers are all set as one. Others are set to zero.
- Hourly Dry Bulb Temperature
- Hourly Wet Bulb Temperature
- Hourly Dew Point Temperature
- Hourly Relative Humidity
- Hourly Wind Speed
- Hourly Wind Gust Speed
- Hourly Station Pressure

The following airport hourly departure and arrival flights data are extracted for the origin and destination airports from the ASPM database.

- Scheduled departures
- Scheduled arrivals

Moreover, there are some Non-numerical features: To analyze the ORD related airports in the research, we used the one-hot encoding by converting the name of airport into numerical features. There are 30 relevant airports selected based on the total number of the departures and arrivals to ORD. The ORD related airport network is shown in Figure 4. The relevant airports are converted to 30 airport features and 1 direction features.

- Direction: 0 is from ORD to other airports; 1 is from other airports to ORD.
- Airport (30): 1 if this airport is active; 0 otherwise.

All BTS, LCD and ASPM data of the 30 ORD related airports are matched for the same flight according to the time. The missing data of BTS is deleted directly. While the data in LCD and ASPM is provided according to time. The missing data in LCD and ASPM is filled by linear interpolation.

C. Oversampling for imbalanced data

The portion of on-time flight over delayed flight is about 80% to 20%, which is not equally represented. The imbalanced data in air delay prediction will cause the machine-learning model to predict biased results when dealing with the delayed data. To make the model to be unbiased, we use the synthetic minority over-sampling technique (SMOTE) method to deal with the imbalanced data [16]. This algorithm generates the synthetic samples from the minority samples. This technique can improve the performance of the learning process effectively. The final proportion of delayed flights are 50%, which is equal to the on-time flights.

IV. Feature selection

Since the effect of the curse of dimensionality, training all the features from BTS, NOAA, ASPM may not have the best performance [11]. The classifiers performance may decrease if the dimension increases without enough training samples. Too many features may confuse the classifier and make it less likely to find the key elements for classification. Normally, feature selection is applied to a portion of samples from the training set to improve the performance. Instead, to pursue the best learning result, we select the best features for all the database with 5-fold cross validation [13].

It is reported that the origin-destination (OD) airports pair may have considerable effects on the flight delay [5]. To consider the impacts of the related airports, the features for airports are added apart from the three databases (BTS, LCD, ASPM). Based on the ORD related airports network, 30 features of airports and 1 direction feature are applied to our model as the OD pair feature. The direction feature is set to be 0 or 1 for ORD as origin or destination. For example, if the flight is from ORD to LGA, then the direction feature is 0, LGA feature is 1 and all the other 29 airport features are set to be zero. The introduction of the direction feature decreases the dimensionality from 60 to 31.

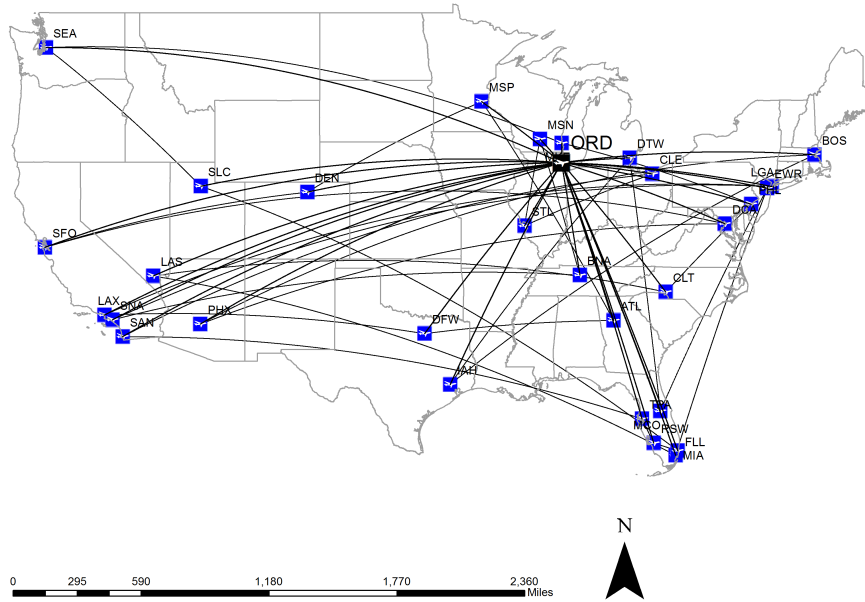


Figure 4. ORD related airports network.

A. Metrics of accuracy

In the past research, accuracy is an important criterion to evaluate the binary classification problem: delay or on-time. Accuracy is defined on a 2-by-2 confusion matrix and it returns the correct prediction portion of all the classifications.

As it is discussed in section III, this paper uses multiple delay groups to represent delays. The delays are strictly divided into 15 categories from -2 to 12. According to the definition of delay group from Federal Aviation Administration, the relationship between the group error and time error is not linear and there is an overlapping time zone. For example, if the predicted delay group is the same with the actual delay group, it means the error gap of predicted delay time and actual delay time is between $[0, 15]$ minutes. If the predicted group index and actual group index is different for one index, it means the error gap in time is between $[1, 30]$ minutes. Only using the traditional accuracy may neglect the overlapping time zone between $[1, 15]$ minutes for one delay group index error, which is not very accurate to evaluate the delay phenomenon.

To include the one delay group index error, we relaxed the prediction error standard slightly and raised a new criterion, called relaxed accuracy for our problem. As it is shown in Figure 5, all the non-positive delay group are computed as on-time zone and reclassified as 0 (left top yellow square). The red squares with zero delay group index error out of the yellow square are the accuracy. The orange squares with one delay group index error are computed in the relaxed accuracy to include the error time in the same delay group. Although these orange zones may increase the delay time error with the range of $[16, 30]$ out of the on-time group, we think this error are acceptable. Relaxed accuracy will not affect the on-time performance and reflect the actual prediction performance more practically. Relaxed accuracy is the sum of the red, orange and yellow zones in the confusion matrix.

B. Recursive feature elimination

A recursive feature elimination (RFE) algorithm is implemented for our feature selection process [17]. As it is shown in Algorithm 1, RFE starts from computing with all the features and generates the feature importance in random forest. Then the least important feature will be eliminated from the feature set. The performance of elimination will be compared in the test data set with cross validation. This procedure will be repeated until the accuracy of the model reaches the highest performance. This algorithm is recursive and may reach local optimized features.

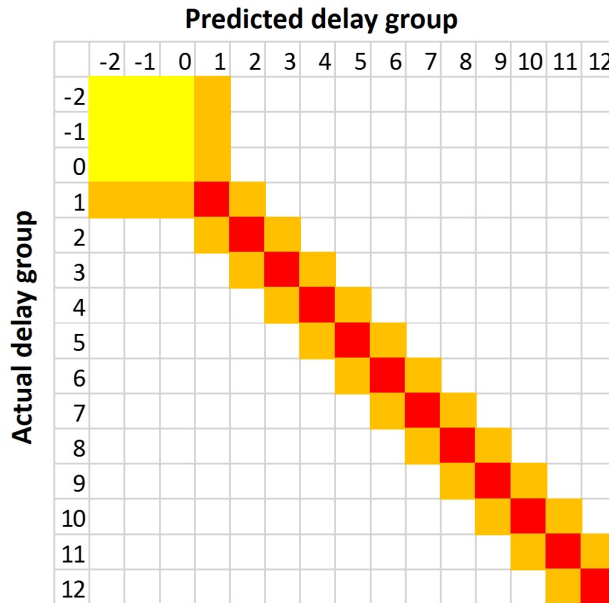


Figure 5. Delay group confusion matrix.

Algorithm 1 Recursive feature elimination

- 1: Initial: training set \mathbb{T}
 set of all features \mathbb{F}
 - 2: For iteration i in $|\mathbb{F}|$:
 compute feature importance with Random Forest
 rank the feature set
 find the last ranked feature f^* in \mathbb{F}
 - 3: Compute the performance with $\mathbb{F} - f^*$
 if performance is improved:
 update: $\mathbb{F} = \mathbb{F} - f^*$
 go back to Step 2 with new \mathbb{F}
 else: break
-

C. Selected features

Random forest can export the feature importance during the building of decision trees, which is a good standard for feature selection. By analyzing the feature importance of arrival and departure prediction individually, some phenomenon corresponding to the flight delay are revealed with the key factors, which could provide practical information for the delay prediction.

The normalized feature importance for arrival delay prediction is shown in Figure 6. For the arrival delay prediction, departure delay importance is much higher than the other features. This is identical to the common sense of the delay propagation from origin to destination. Figure 7 demonstrates the optimal features after feature selection. All the other features hold mostly the same importance. Most of the weather features and half of the BTS features are filtered in the selection. All the ASPM features are kept in the final feature set. These indicates that most of the weather factors, such as the precipitation and wind gust speed have little impact on the commercial airlines. The operation condition of the origin and destination airports are important for the flights. Therefore, the airport congestion conditions effect on the delay cannot be neglected. Compared to the features from three main databases, the airport features have little importance to the delay. The delay mechanism is almost the same for most of the OD pairs in arrival delay prediction and for most of the airports when predicting the departure delay. These proves the necessary of feature selection and the introduction of the ASPM database to consider the terminal congestion.

Figure 8 and 9 shows the characterizing of departure delay. For the departure delay prediction, LAAD is much higher than the other features. Although the LAAD is still the most important feature for departure delay prediction, it doesn't dominate the feature importance chart like the departure delay importance does in arrival delay prediction.

Arrival delay feature importance

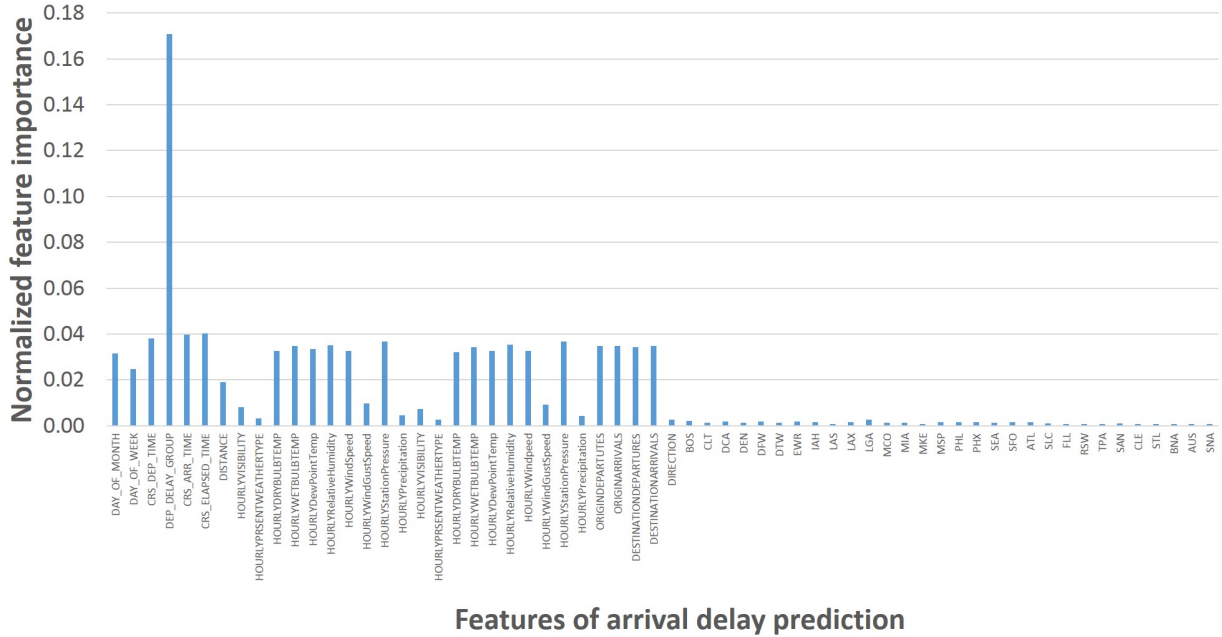


Figure 6. Normalized feature importance for arrival delay prediction

Feature importance of arrival delay

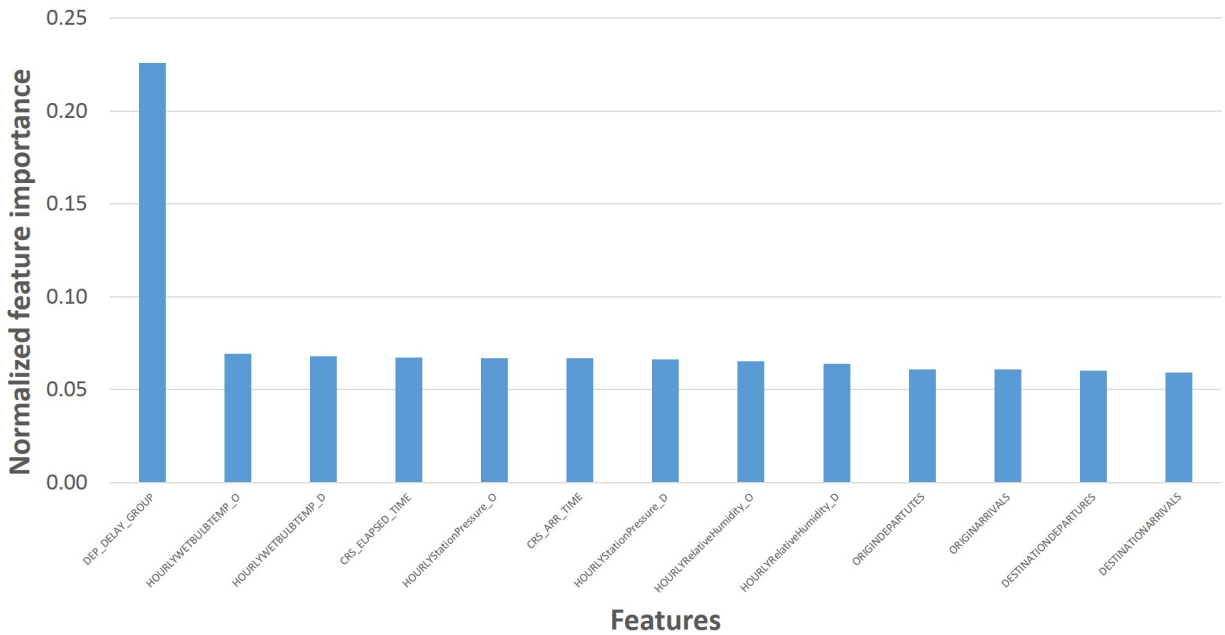


Figure 7. Optimal feature importance for arrival delay prediction

Prediction through the air route is different from the prediction within the airports. This difference is caused by several reasons. First, only LAAD, which is the only computed cause of the five delay causes, is kept for the model. Although

all the five delay causes are used to explain the arrival delay, only LAAD can be acquired before the flights take off and is directly related to the departure delay. Deficiency of the other four features decreases the prediction accuracy, but in the actual prediction they are difficult to be acquired before the flight taking off. Second, LAAD is actually a human explained data. As the carriers have to satisfy the requirement that the sum of the five delay causes time should be equal to the final arrival delay, LAAD may be inconsistent with the actual LAAD according to its definition. This difference between LAAD and actual LAAD will be proved in the later section. Third, departure delay prediction in the airport is more chaotic than the arrival delay prediction. Carriers prefer ground delays in the airport rather than airborne delays. The airport congestion and schedule will have other effects to the departure delay.

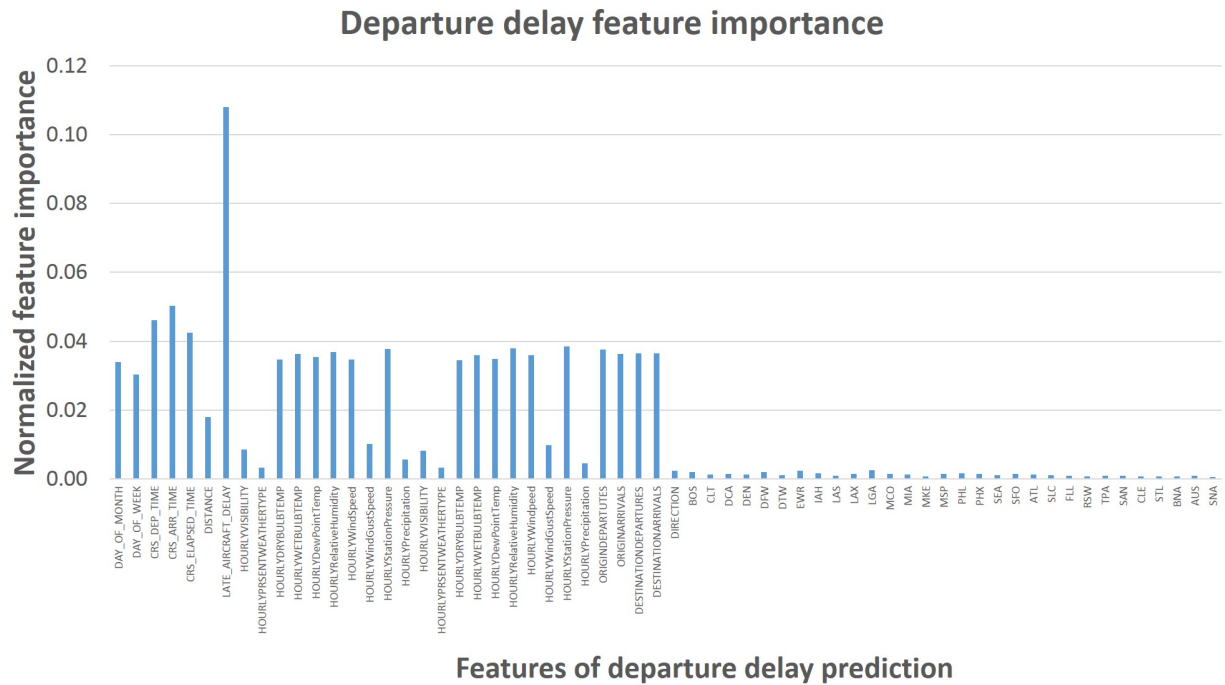


Figure 8. Normalized feature importance for departure delay prediction.

To analyze the effect of airports, the importance of airports is shown in the Figure 10. Airport features include 1 direction feature and 30 ORD related airport features. These features are of little importance compare to the other features in Figure 8. Most of machine learning based air traffic systems would focus on some particular OD pairs or airports. But this chart provides a credential that difference of the links can be abandoned if the dataset is not sufficient to support the classifier. Building the independent classifiers for each particular links may be less worthwhile than expected. Since combination of direction and airport features is used to represent 1 OD pair, the direction feature is the summary of the 30 OD pairs. The high importance of the directions shows that different directions of the two airports have different impacts on delay.

Figure 10 also proves the phenomenon that the more congested the airport is, the more importance of the airport feature is. Compare to the main airports of the air traffic systems, small airports are less specified and can be unified as one airport. The features of the small airports may decrease the performance of the classifier. The departure delay and the arrival delay importance are consistent for most of the airports, which proves that the airport effects on the two kinds of delay are similar.

Figure 11 and 12 provide the error distribution along the error group for both predictions with all features and optimal features. Recursive feature selection improves the performance of the model in both accuracy and relaxed accuracy, as it is shown in Table 2. The improvement of arrival delay is more notable than the departure delay prediction, which is because of the complexity of the departure delay’s causes. But in both cases of prediction group 12, the improvement does not make the difference. This is partly caused by the max limitation of the delay group. The above prediction results prove that adding all the features for the classifier will confuse the learning model. We used the selected optimal features in the later iteration model application.

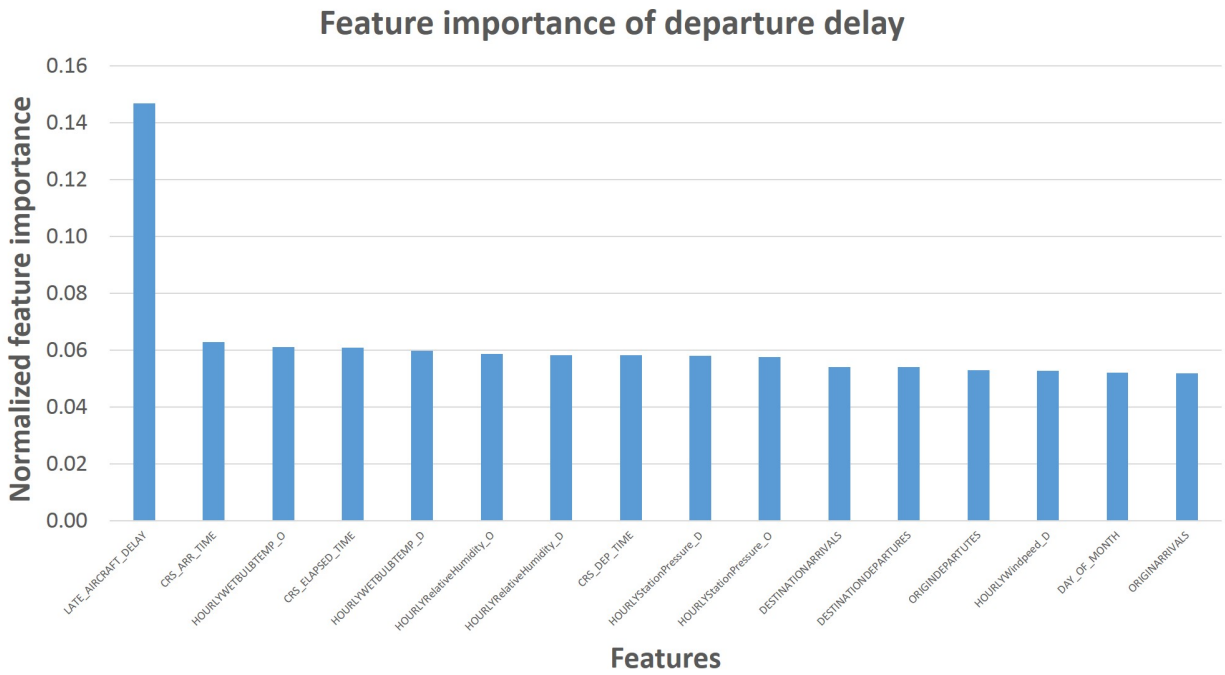


Figure 9. Optimal feature importance for departure delay prediction

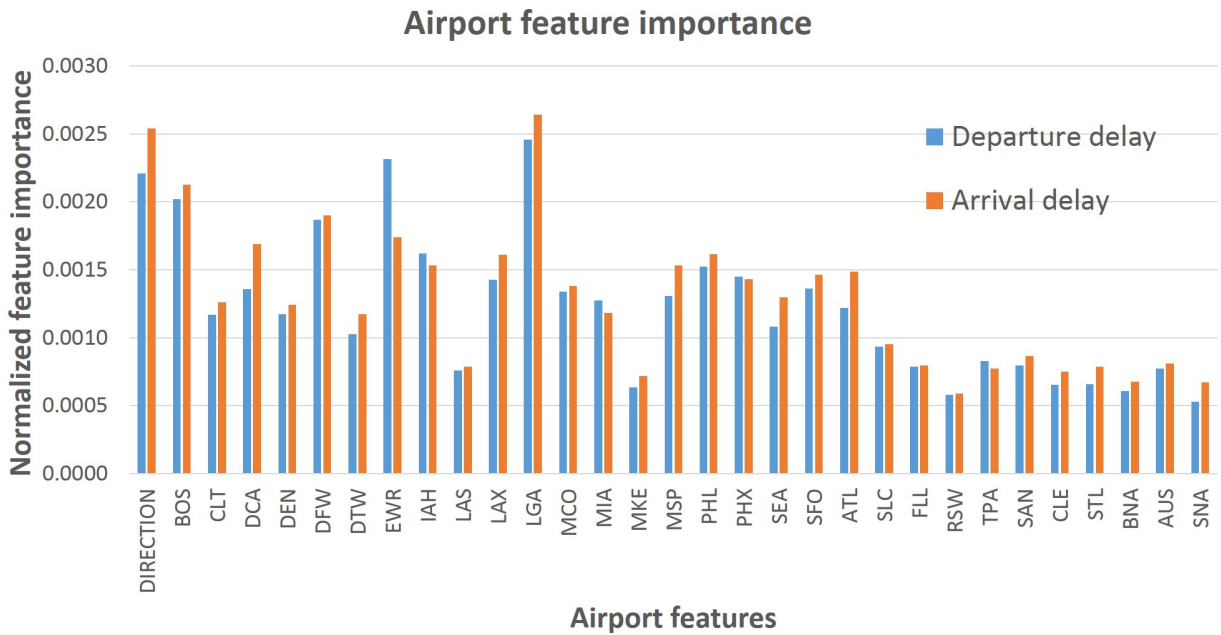


Figure 10. Airport feature importance for delay prediction

Table 2. Comparison of the accuracy

	Arrival Delay all features	Arrival Delay optimal features	Departure Delay all features	Departure Delay optimal features
Relaxed Accuracy	0.87498	0.92669	0.88244	0.90354
Accuracy	0.85915	0.86727	0.82684	0.83050

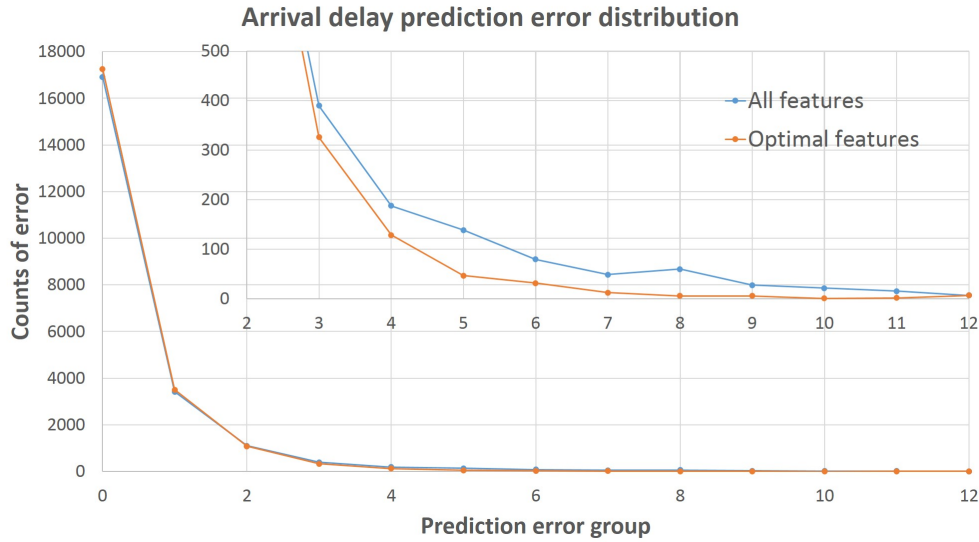


Figure 11. Arrival delay error distribution with error group

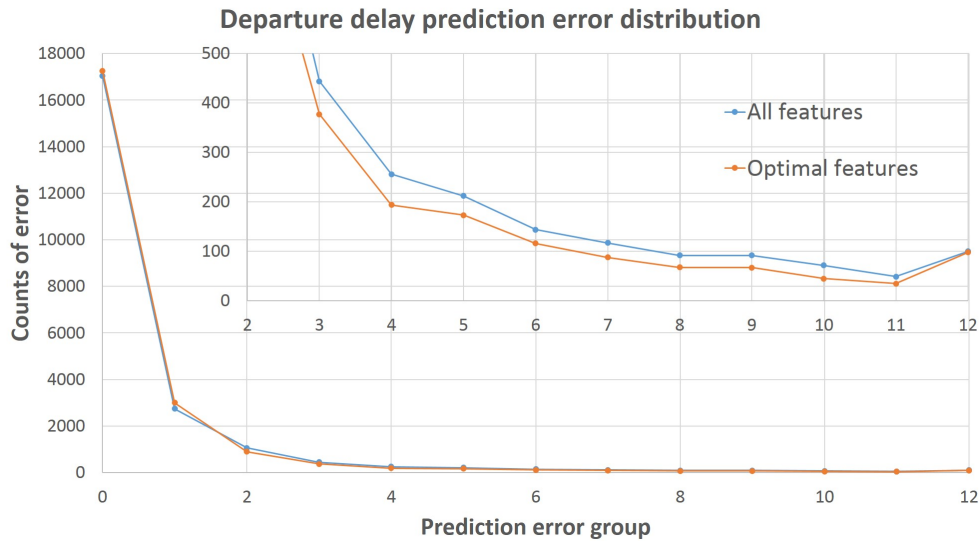


Figure 12. Departure delay error distribution with error group

V. Application

In the chained delay prediction model, predictions of the departure delay and arrival delay are achieved directly by random forest classification. The transition from the arrival delay to LAAD can be considered as the delay propagation within the airports. The traditional airline arrival delay prediction requires the actual departure delay time, which is a

strong condition and can only predict one time of the flight. Introducing the delay propagation can increase the time range of the prediction by working as a chained model. Although the performance of the chained model will decrease as the iteration number increases, this disadvantage can be fixed by dynamically updating the actual departure delay. An initial departure delay should be given in the beginning of the daily operation as the initial input for the chained model to work.

A. Chained delay prediction with ORD-Related airports

To demonstrate the performance of our chained delay prediction model, we select the ORD related flights for application. The flights with ORD as origin or destination are defined as the ORD related flights. The flights operated in January 2017 within the ORD related airports network (see Figure 4) are chosen as the test set. There are 26693 ORD related flights satisfying the required condition.

In the set of ORD related flights, all the aircraft will arrive in or departure from ORD in every flight. Every aircraft will follow its scheduled order of itinerary in the traffic system. The order of the same aircraft during a day is defined as the iteration number. The iteration number reflects the times the flight is operated in the air traffic system of the day. The iteration number is also the key parameter in the model of the air traffic network. The total flights distribution with the iteration number is shown in Figure 13. In the ORD related air traffic system, the maximum iteration number is 10, and 97.52% of the flights have less than 6 iteration numbers.

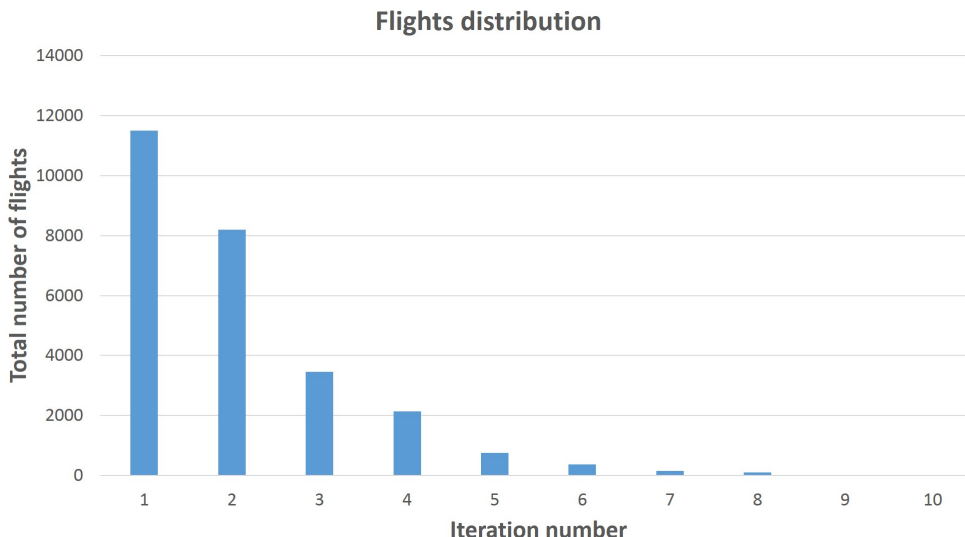


Figure 13. Flights distribution with iteration number

In order to test the effect of the delay propagation, we defined three different test models as following:

- CH10: Chained prediction model with all actual departure delays for up to iteration number 10.
- CH1: Chained prediction model with initial actual departure delay and delay propagation
- CH1-NPRG: Chained prediction model with initial actual departure delay and without delay propagation

CH10 predicts all the flights behavior along the itinerary with actual departure delay. This model represents the best performance of random forest classification, which is the upper bound of the chained model. CH1 runs with the initial departure delay to predict the first arrival delay and then predicts the following flights with delay propagation. CH-NPRG runs with initial departure delay without the delay propagation module, which means all the following departures and arrivals are exactly right on time and will not be affected by actual data. In other words, the LAAD feature for departure prediction and the departure delay feature for arrival delay prediction are both set to zero. CH1-NPRG is a contrast to CH1 to show the performance of delay propagation. The comparison of three models can reflect the impact of the delay propagation by evaluating the total counts of prediction error.

As it is shown in Figure 14, the three models have similar distribution from error group 0 to 12. But it still can be noticed that this is a little difference in the region with error group 0. This is caused by the accumulation of error in

region of error group 2 to 12. From the amplified region in right upper corner of Figure 14, the difference is notable. As it is expected, model CH10, which is the complete classification with actual departure time of all the iteration numbers, has the lowest error and highest accuracy. Model CH10 is also the upper bound of the prediction accuracy. By comparing the model CH1 and CH1-NPRG, we can found that adding the effect of the delay propagation will improve the performance of the prediction by decreasing the error in region of error group 2 to 12.

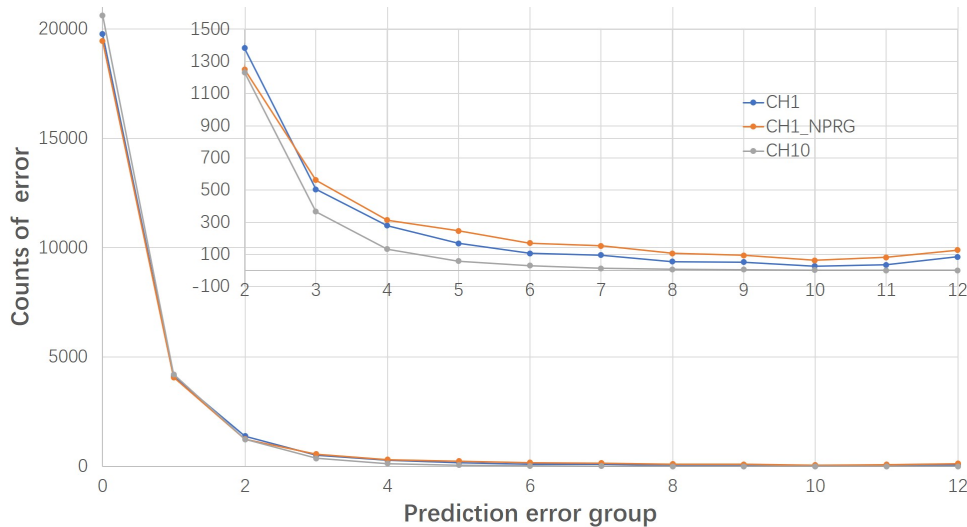


Figure 14. Performance comparison

Although the effect of delay propagation is checked here, it can still be observed that the difference of the CH1 and CH10 is larger than the difference between CH1 and CH1-NPRG. This is because only the initial input is actual and correct. In other words, only the flights of iteration number 1 are updated into the model.

B. Impact of the iteration number

To check the effect of the iteration number of the flights in chained delay model, we develop the definition of the CH to Chi, where “i” represents how many previous actual departure delays are updated into the chained delay model. Iteration number i is from 1 to 10. By controlling the iteration number of the flights data updated into the model, the impact of the iteration number can be revealed through the simulation result. Two criterions (i.e. relaxed accuracy and accuracy) are set to analyze the model with the iteration number. The distribution of relaxed accuracy and accuracy with the iteration number are shown in 15.

As the actual departure delay are updated into the model, both the accuracy and relaxed accuracy increase with the iteration number. Since the total number of data points is constant, the data points in relaxed accuracy region (i.e. error group 0 and 1) will increase and the ones with error group 2 to 10 will decrease. The higher the prediction error is, the faster the total number decreases with the iteration number. The convergence speed decreases as the iteration number increases. All the criterions converge as the iteration number increases.

To give a more directly conclusion of the prediction error distribution, normalized prediction error distribution with iteration number and error class is shown in Figure 16. Since the number of error counts of high error class is too small to observe if compared with that of lower error class, all the prediction errors are normalized by the corresponding initial prediction error of CH1 to observe the distribution of high error class more clearly. The error class of 0 and 1, which is also the relaxed accuracy region, will increase slightly as the iteration number increases. On the contrary, the other class will decrease with the iteration number. The higher the error class is, the faster will the prediction error decrease. This also proves that by updating the actual departure time, the prediction will be more and more accurate and the prediction error will decrease.

VI. Conclusions

This paper presented a new machine learning based air traffic delay prediction model that combined multi-label random forest classification and approximated delay propagation model. To improve the prediction performance, an optimal feature selection process is introduced and demonstrated to have better performance than directly using all the features

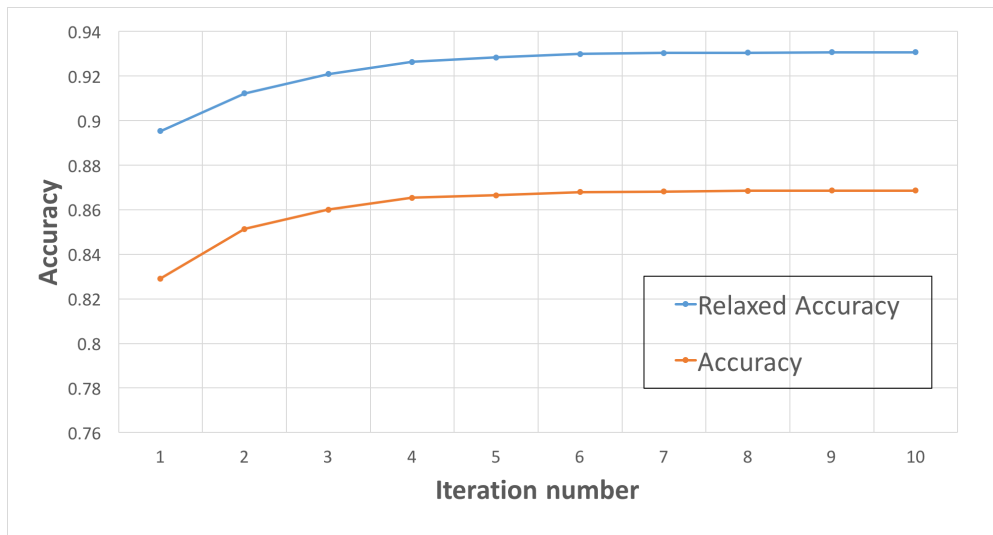


Figure 15. Accuracy and relaxed accuracy with the iteration number

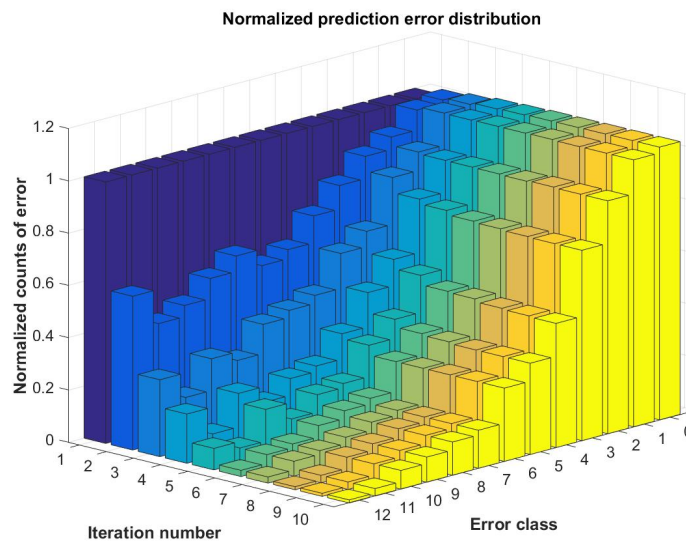


Figure 16. Normalized prediction error distribution

of available datasets. Departure delay and late arriving aircraft delay are shown to be the most important features for delay prediction. To utilize these two features, a delay propagation model is proposed as a link to connect them to build a chained delay prediction model. Given the initial departure delay, the chained model is demonstrated to have the ability to predict the flight delay along the same aircraft’s itinerary. By updating the actual departure delay with the iteration number along the itinerary, the models accuracy can be further improved. This machine learning based method has been shown to be more accurate and practical for delay prediction in daily air traffic operation.

References

[1] Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A. A., and Zou, B., “Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States, NEXTOR,” 2010.

[2] Chen, J. and Sun, D., “Stochastic Ground-Delay-Program Planning in a Metroplex,” *Journal of Guidance, Control, and Dynamics*, Vol. 41, No. 1, 2017, pp. 231–239.

- [3] Ferguson, J., Kara, A. Q., Hoffman, K., and Sherry, L., "Estimating domestic US airline cost of delay based on European model," *Transportation Research Part C: Emerging Technologies*, Vol. 33, 2013, pp. 311–323.
- [4] Chen, J., Chen, L., and Sun, D., "Air traffic flow management under uncertainty using chance-constrained optimization," *Transportation Research Part B: Methodological*, Vol. 102, 2017, pp. 124–141.
- [5] Pyrgiotis, N., Malone, K. M., and Odoni, A., "Modelling delay propagation within an airport network," *Transportation Research Part C: Emerging Technologies*, Vol. 27, 2013, pp. 60–75.
- [6] Chen, J., DeLaurentis, D., and Sun, D., "Dynamic stochastic model for converging inbound air traffic," *Journal of Guidance, Control, and Dynamics*, 2015, pp. 2273–2283.
- [7] Mueller, E. R. and Chatterji, G. B., "Analysis of aircraft arrival and departure delay characteristics," *AIAA aircraft technology, integration and operations (ATIO) conference*, 2002.
- [8] Klein, A., "Airport delay prediction using weather-impacted traffic index (WITI) model," *Digital Avionics Systems Conference (DASC), 2010 IEEE/AIAA 29th*, IEEE, 2010, pp. 2–B.
- [9] Xu, N., Donohue, G., Laskey, K. B., and Chen, C.-H., "Estimation of delay propagation in the national aviation system using Bayesian networks," *6th USA/Europe Air Traffic Management Research and Development Seminar*, 2005.
- [10] Rebollo, J. J. and Balakrishnan, H., "Characterization and prediction of air traffic delays," *Transportation research part C: Emerging technologies*, Vol. 44, 2014, pp. 231–241.
- [11] Choi, S., Kim, Y. J., Briceno, S., and Mavris, D., "Prediction of weather-induced airline delays based on machine learning algorithms," *Digital Avionics Systems Conference (DASC), 2016 IEEE/AIAA 35th*, IEEE, 2016, pp. 1–6.
- [12] Breiman, L., "Random forests," *Machine learning*, Vol. 45, No. 1, 2001, pp. 5–32.
- [13] Hastie, T., Tibshirani, R., and Friedman, J., "Overview of supervised learning," *The elements of statistical learning*, Springer, 2009, pp. 9–41.
- [14] Ho, T. K., "The random subspace method for constructing decision forests," *IEEE transactions on pattern analysis and machine intelligence*, Vol. 20, No. 8, 1998, pp. 832–844.
- [15] Chen, J., Cao, Y., and Sun, D., "Modeling, Optimization, and Operation of Large-Scale Air Traffic Flow Management on Spark," *Journal of Aerospace Information Systems*, 2017, pp. 504–516.
- [16] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P., "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, Vol. 16, 2002, pp. 321–357.
- [17] Granitto, P. M., Furlanello, C., Biasioli, F., and Gasperi, F., "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemometrics and Intelligent Laboratory Systems*, Vol. 83, No. 2, 2006, pp. 83–90.