Understanding and Monitoring Different Types of Flight Events Through Machine Learning Classification

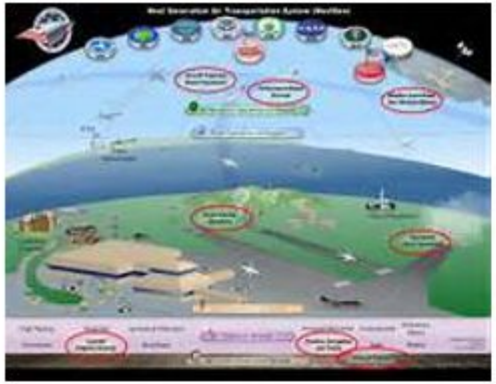
Yutian Pang

Outline

- Introduction
- Methodology
- Results
- Conclusions and Future work

Problem

Safe &
Effective NAS



- Large amount of ATM data stay unexplored in the database.
- Recent technology advances in sensors, networking, data mining and other analytic techniques enable proactive risk management for National Airspace System(NextGen)
- Machine learning tends to be a tool that can help understanding the huge data.

Challenge

How to classify the aircraft event given different recorded features?

lKey	cKey	SysName	StartDate	StartTime	tMidnight	tStartSecs	tStopSecs	tStart	tStop	Duration	Msn	AcId	AcType	tEv	EvType	ObjClass	OldName	NewName	Lat	Lon	aEv	cEv	vEv	rEv	DTD	FlD	DDT	
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	80436	EV_TOF	APT	?	JFK	40.62412	-73.7843	14	206	208.5	0	0	0		
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	80436	EV_USER	EVT	?	TOF	40.62412	-73.7843	14	206	208.5	0	0	0	0	
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	80436	EV_USER	SEG	TOF	LND	40.62412	-73.7843	14	206	208.5	0	2221.502	2294.478	72.976	22386
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	80436	EV_USER	SEG	TOF	inARTCC	40.62412	-73.7843	14	206	208.5	0	46.141	84.825	38.684	1039
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	80455	EV_MOF	VER	UNK	CLB	40.61167	-73.7922	14	206	208.5	0	0	0	0	0
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	80455	EV_INIT	CTR	OUTSIDE	OUTSIDE	40.61167	-73.7922	14	206	208.5	0	0	0	0	0
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	80455	EV_INIT	SCT	NNN@ZN	NNN@ZN	40.61167	-73.7922	14	206	208.5	0	0	0	0	0
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	80455	EV_TRNS	REQALT	UNK	BLW	40.61167	-73.7922	14	206	208.5	0	0	0	0	0
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	81066	EV_MOF	VER	CLB	LVL	40.98111	-73.7953	170	283	328.5	628	21.983	48.021	26.037	611
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	81475	EV_XING	CTR	OUTSIDE	ZNY@ZNY	41.11625	-74.5767	170	283	327.7	0	46.343	84.001	37.658	1020
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	81475	EV_XING	SCT	NNN@ZN	35@ZNY	41.11625	-74.5767	170	283	327.7	0	46.343	84.001	37.658	1020
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	81475	EV_USER	EVT	?	inARTCC	41.11625	-74.5767	170	283	327.7	0	0	0	0	0
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	81475	EV_USER	SEG	inARTCC	outARTCC	41.11625	-74.5767	170	283	327.7	0	2123.293	2145.775	22.482	20399.98
14096	USA_2020	usa	1/6/2020	22:20:36	1.58E+09	1.58E+09	1.58E+09	80436	104754	24318	29479	AAL2652	A321	81493	EV_MOF	VER	LVL	CLB	41.12167	-74.6089	170	283	332	154	37.46	37.461	0.001	427

Flight Event Data from NASA Sherlock Data Warehouse

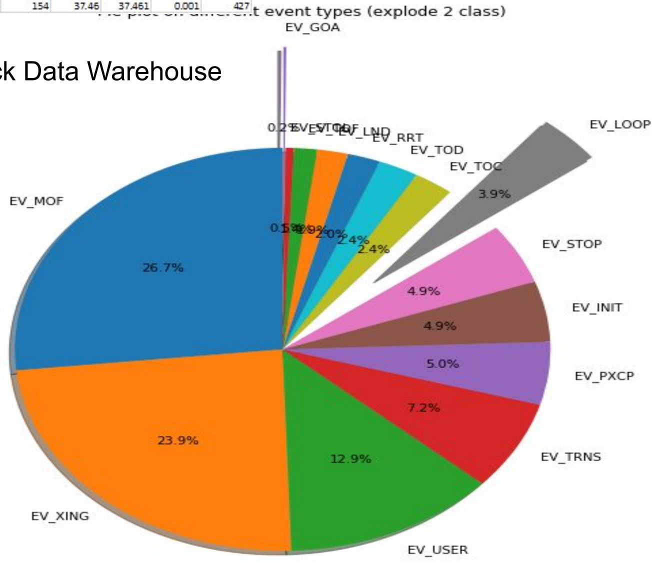


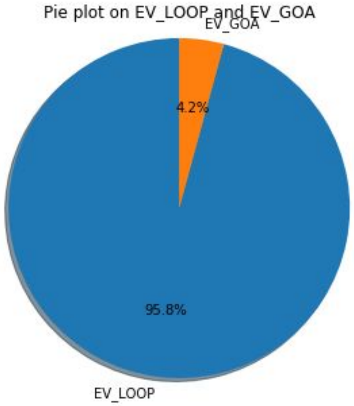
Table 1 Prediction Labels

Event Type	Counts	Meaning
EV_MOF	279645	Mode of flight. A flight record in vertical domain such as descending level, climbing level. Also known as vertical trajectory.
EV_XING	250173	Crossing event from one sector to another sector. Indicate which space volume are you in and which one are you crossing to other column.
EV_USER	134956	User event. The definition is flexible includes times you go into a center, different volume definitions. It's a user based concept.
EV_TRNS	75086	Transition from above or below altitude. Not often used.
EV_PXCP	52892	Unknown.
EV_INIT	50934	Detail on the begin of flight tracks. Smaller set on facility recorded by the surveillance system.
EV_STOP	50934	The last track point of the aircraft.
EV_LOOP	41118	Indication of holding pattern. The flight circle around in the trajectory.
EV_TOC	25467	Top of climb. Reach cruise altitude. Overlap with MOF.
EV_TOD	25467	Top of descend. Start to descent from cruise altitude
EV_RRT	20692	Reroute. New flight plan change compared to old flight plan.
EV_LND	19526	Landing event. Recorded when arrival aircraft pass the arrival runway threshold.
EV_TOF	14697	Take off event. It's defined geometrically when the aircraft cross the departure runway threshold (wheels come off the pavement).
EV_STOL	5174	Stop holding. Finish EV_LOOP Event.
EV_GOA	1814	Go around. About to land but pull up and circle back, turn around again. A possible safety concern.

Challenge

Table 1 Prediction Labels

Event Type	Counts	Meaning
EV_MOF	279645	Mode of flight. A flight record in vertical domain such as descending level, climbing level. Also known as vertical trajectory.
EV_XING	250173	Crossing event from one sector to another sector. Indicate which space volume are you in and which one are you crossing to other column.
EV_USER	134956	User event. The definition is flexible includes times you go into a center, different volume definitions. It's a user based concept.
EV_TRNS	75086	Transition from above or below altitude. Not often used.
EV_PXCP	52892	Unknown.
EV_INIT	50934	Detail on the begin of flight tracks. Smaller set on facility recorded by the surveillance system.
EV_STOP	50934	The last track point of the aircraft.
EV_LOOP	41118	Indication of holding pattern. The flight circle around in the trajectory.
EV_TOC	25467	Top of climb. Reach cruise altitude. Overlap with MOF.
EV_TOD	25467	Top of descend. Start to descent from cruise altitude
EV_RRT	20692	Reroute. New flight plan change compared to old flight plan.
EV_LND	19526	Landing event. Recorded when arrival aircraft pass the arrival runway threshold.
EV_TOF	14697	Take off event. It's defined geometrically when the aircraft cross the departure runway threshold (wheels come off the pavement).
EV_STOL	5174	Stop holding. Finish EV_LOOP Event.
EV_GOA	1814	Go around. About to land but pull up and circle back, turn around again. A possible safety concern.



- Out of all the aircraft events, engineers are interested in EV_LOOP (loop) and EV_GOA (go around).
- The classes are extremely imbalanced.

Outline

- Introduction
- Methodology
- Results
- Conclusions and Future work

Machine learning based classification

- Preprocessing
  - Documented features
  - Generated features
- Handle Imbalanced data
  - Undersampling
  - Oversampling
  - Combination of undersampling and oversampling
- Classification
  - Logistic regression
  - SVC
  - Naive Bayes
  - K neighbor classifier
  - Decision tree
  - Random Forest
  - Gradient Boosting

Algorithm 1 Airspace Density Feature

Data: raw flight event file
Input: $\zeta_{time}, \zeta_{lon}, \zeta_{lat}$
Output: df$_2$

Load row flight event file in df
$df_2 \leftarrow df['EvType'] =' EV\_LOOP' \cup df['EvType'] =' EV\_GOA'$
Create empty column $'Count'$ in df_2
for index, row in df_2 **do**
    $time_{idx}$ ←abs(df['tEv'] - row['tEv']) < ζ_{time}
    lat_{idx} ←abs(df['Lat'] - row['Lat']) < ζ_{lat}
    lon_{idx} ←abs(df['Lon'] - row['Lon']) < ζ_{lon}
    tol_{idx} ←time$_{idx}$ ∩lat$_{idx}$ ∩lon$_{idx}$
    $df_2[index,'Count']$ ←df$_2[total_idx]['AcId'].unique().size$
end for

Algorithm 2 Precious Event Sequence

Data: df_2
Output: df_2

Get dummie variables on $EvType$
index_old← 0
for index, row in df_2 **do**
    Get dummie variable on $df[index\_old : index - 1]['EvType']$
    Sum along rows of $df[index\_old : index - 1]$
    $df_2[index] \leftarrow appendrow$
    $index\_old \leftarrow index$
end for

Outline

- Introduction
- Methodology
- Results
- Conclusions and Future work

Results

Classifier	Data Balance	Precision	Recall	F1-score	Accuracy
Logistic Regression	imbalanced	0.61	0.61	0.61	0.94
	balanced	0.91	0.91	0.91	0.96
Support Vectors	imbalanced	0.61	0.61	0.61	0.94
	balanced	0.88	0.87	0.87	0.88
Naive Bayes	imbalanced	0.61	0.61	0.61	0.94
	balanced	0.81	0.72	0.71	0.74
KNN	imbalanced	0.61	0.61	0.61	0.94
	balanced	0.99	0.98	0.98	0.98
Decision Tree	imbalanced	0.54	0.67	0.52	0.79
	balanced	0.98	0.97	0.97	0.97
Random Forest	imbalanced	0.61	0.61	0.61	0.94
	balanced	**0.99**	**0.99**	**0.99**	**0.99**
Boosting	imbalanced	0.61	0.61	0.61	0.94
	balanced	0.96	0.96	0.96	0.96

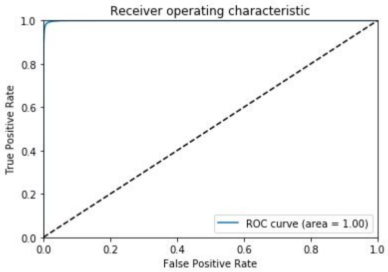
Random Forest Both Trained with Balanced data

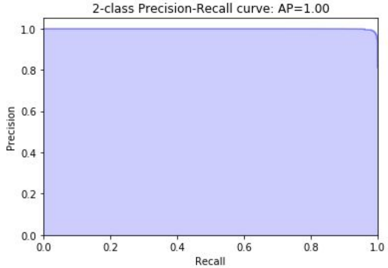
```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99     13189
           1       0.99      0.99      0.99     11488

    accuracy                           0.99     24677
   macro avg       0.99      0.99      0.99     24677
weighted avg       0.99      0.99      0.99     24677

[[13078   111]
 [  167 11321]]
Area under the ROC curve (AUC): 0.999382
```





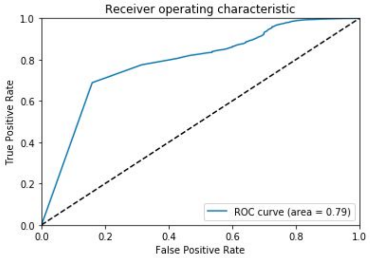
Tested with Balanced data

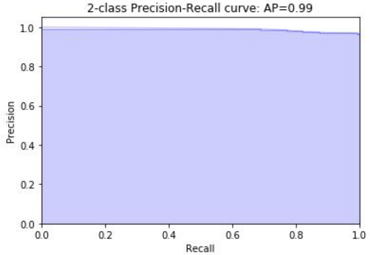
```
              precision    recall  f1-score   support

           0       0.28      0.24      0.26       546
           1       0.97      0.98      0.97     13622

    accuracy                           0.95     14168
   macro avg       0.63      0.61      0.62     14168
weighted avg       0.94      0.95      0.95     14168

[[  129   417]
 [  326 13296]]
Area under the ROC curve (AUC): 0.786197
```





Tested with Imbalanced data

Feature Importance

Table 3 Feature Importance of RandomForestClassifier

Rank	Feature	Description
1	FlD	flight distance since previous event of same category
2	EvType_EV_MOF	mode of flight eventin vertical domain
3	FlT	flight time since previous event of same category
4	DDT	deviation from direct-to since previous event of same category
5	EvNumInfo	event number - additional placeholder for any integer event information.

Outline

- Introduction
- Methodology
- Results
- Conclusions and Future work

Conclusions and future work

- Conclusions,
  - The RF model is able to perfectly classify two flight events with preprocessed imbalanced data.
  - The performance is bad when applying to unprocessed dataset.
- Future directions,
  - Better approach to handle imbalanced data
  - Multiclass classification
  - Non-deterministic classification models

References

[1] Lee, H., Li, G., Rai, A., and Chattopadhyay, A., "Anomaly detection of aircraft system using kernel-based learning algorithm," *AIAA Scitech 2019 Forum*, American Institute of Aeronautics and Astronautics Inc, AIAA, 2019. doi:10.2514/6.2019-1224.

[2] Liu, Y., and Goebel, K., "Information Fusion for National Airspace System Prognostics," *PHM Society Conference*, Vol. 10, No. 1, 2018, pp. 1–13.