Analysing Incidents using Data Enrichment and Machine Learning

Gerard van Es

ICAO First iSTARS User Group (iUG/01)
ICAO Headquarters, Montreal – 17-19 December 2018
Background – occurrence data

Occurrence data

Operators

Ground services

ATC

Airport
Background – Issues with reported occurrences

- Incomplete reports (data fields missing);
- Incorrect reports (errors in data fields).

Limits our data analyses!
How to solve these problems?

**Airport databases**

**Occurrence data**

**Register database**

**METARs**

**Radar/ADS-B**
Example using datamining to code occurrence category

- Random Forest classifier (gave best results out of 5 techniques);
- Random Forest builds multiple decision trees and merges them together to get a more accurate prediction;
- Performance on training dataset gave accuracy of 94%;
- Performance on complete set gave accuracy of 79% (+48,000 occurrences);
- Difficulties with SCF-NP/PP and ADRM & ATC occurrences.
Occurrence data enrichment

- Add information to occurrence data normally not reported, e.g.:
  - Flown track;
  - Duration of flight;
  - Additional weather data (en-route);
  - Taxi track;
  - Non-occurrence flights with same details as for occurrences.

Extends analyses possibilities
Why look at non-occurrence flights?

- Normalising of occurrences (classical approach);
- Allows to explore contribution of certain factors to overall risk increase;
- Will highlight importance of factors not visible from occurrence data only.
Example case

- Reported go-arounds at an airport;
- What is driving the rate of go-arounds at the airport?
- Simply plotting number of go-arounds by year etc. will not be very meaningful...
Typical reasons for a go-around

- Blocked runway, 21%
- Wind (cross of tail), 14%
- Unstabilised approach, 12%
- Separation, 10%
- Landing gear problems, 10%
- Windshear, 6%
- Poor visibility, 5%
- Heavy rain, 2%
- GPWS warning, 1%
- Other, 19%

How important are these factors?
Relation between crosswind and go-around rate
go-around risk ratio in crosswind condition

Cross Wind [kts]

Risk Ratio

0-5
5-10
10-15
15-20
20-25
25-30

[Bar chart showing risk ratio for different crosswind conditions]
Example Machine Learning / Data Mining methods

* Credits: Innaxis Keynote, EUROCONTROL TIM/ART workshop ML and AI, 24-25 sept 2018
Text mining of occurrence data by applying t-SNE algorithm to find patterns
Applying machine learning to enriched occurrence data

- Will help to predict under which conditions there is higher risk of a particular occurrence;

- For instance, combination of wind, period in time, light conditions and runway use could increase go-around rate;

- Need sufficient amount of occurrence data to build accurate models.
Machine Learning – important rules

- You need sufficient and reliable data!

- Define your outcome!

- It is not a push button exercise – You must understand aviation!
Fully engaged
Netherlands Aerospace Centre

NLR Amsterdam
Anthony Fokkerweg 2
1059 CM Amsterdam
p ) +31 88 511 31 13  f ) +31 88 511 32 10
e ) info@nlr.nl  i ) www.nlr.nl

NLR Marknesse
Voorsterweg 31
8316 PR Marknesse
p ) +31 88 511 44 44  f ) +31 88 511 42 10
e ) info@nlr.nl  i ) www.nlr.nl