IDENTIFYING ANOMALIES
in past en-route Trajectories with Clustering and Anomaly Detection Methods

Xavier Olive, Luis Basora

19 June 2019 @ ATM Seminar, Vienna
Follow-up work based on two SID 2018 publications:

- *Detecting Controllers’ Action in Past Mode S data by Autoencoder-Based Anomaly Detection:*
  
  Detection of ATC actions, deconfliction, sequencing, regulations and weather related anomalies based on autoencoders.

- *Occupancy Peak Estimation from Sector Geometry and Traffic Flow Data:*
  
  Occupancy Peak Estimation based on a DBSCAN-based progressive trajectory clustering to identify sector flows.
Detect operationally significant events in traffic history.

Applications are manyfold:

- performance assessment and predictability;
- safety assessment (collision risk models);
- preparation of realistic simulations;
- ML-based/statistical ATC action prediction
DATA SOURCES

- ADS-B tracks (extended Mode S may bring insight) from OpenSky Network †, first semester of 2017 (14,000+ trajectories)
- Sector Configuration Plans *, opening schemes, kindly provided by LFBB ACC
- Demand Data Repository ‡→* helped analyse context, as well as METAR ‡, satellite-based weather analysis *.

† open access to academics
‡ limited access
* closed access
Previous work was on city pairs, let’s bring it to the sector level.
Dimensionality reduction technique, commonly used in anomaly detection

- 1st step: compression/encoding/projection
- 2nd step: decompression/decoding/generation

Core idea: project data to a lower dimensional space and learn to reconstruct it (minimize the MSE loss)

Past experience:

- high reconstruction error → regulations, weather events
- “end of the tail” → ATC deconfliction and sequencing
We chose a stacked network architecture to try to better grasp the structure in each cluster.

\[(S_i) \xrightarrow{\text{ReLU activation}} \text{Linear}(50, 24) \xrightarrow{\text{ReLU activation}} \text{Linear}(24, 12) \xrightarrow{\text{ReLU activation}} \text{Linear}(12, 24) \xrightarrow{\text{ReLU activation}} \text{Linear}(24, 50) \xrightarrow{\text{sigmoid activation}} (\hat{S}_i)\]
ANOMALY DETECTION PER CLUSTER
Since the distribution of the reconstruction errors looks like it is favouring one “mode” over the other, we designed a regularisation term to favour an exponential distribution:

MSE loss: \[ \ell(u, v) = \frac{1}{n} \sum ||u_i - v_i||^2 \]

fitted distribution: \[ f : x \mapsto \frac{1}{\ell(u, v)} \cdot \frac{x}{e^{-x \ell(u, v)}} \]

difference to minimize: \[ \delta_j = \left( \frac{1}{n} \sum_i 1_{[t_j, t_{j+1}]}(\rho_i) \right) - f(t_j) \]

regularized loss: \[ \ell^*(u, v) = \ell(u, v) + \lambda \sum_{j=1}^m \max(0, \delta_j) \]
Hit “Run”.
### WHO IS THE BAD GUY? (TOP 10 PER CLUSTER)

<table>
<thead>
<tr>
<th>callsign</th>
<th>date of entry</th>
<th>cluster</th>
<th>rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRA47R</td>
<td>2017-06-20 18:21:00Z</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>RYR9TG</td>
<td>2017-06-20 19:09:28Z</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>TOM84T</td>
<td>2017-06-20 19:14:51Z</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>RYR79EY</td>
<td>2017-06-20 18:35:46Z</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>SAA235</td>
<td>2017-06-20 19:07:15Z</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>DAH1007</td>
<td>2017-06-20 18:33:34Z</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>DLH68F</td>
<td>2017-06-20 18:25:17Z</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>VLG83TJ</td>
<td>2017-06-20 18:21:00Z</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>AFR88DM</td>
<td>2017-07-08 19:55:01Z</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>VLG8248</td>
<td>2017-07-08 20:09:53Z</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>AEA1008</td>
<td>2017-07-08 19:39:50Z</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>AAF221</td>
<td>2017-07-08 20:08:52Z</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>FIN611</td>
<td>2017-07-08 19:42:22Z</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>
CO-OCCURRENCES OF DETECTED EVENTS
DIFFERENT FLOWS (FIG. 11)
DIFFERENT FLOWS (FIG. 12)

- **2017-01-03 06:26:14 UTC**: 5 minutes before CPA
  - TVF021Z
  - FL390
  - TAP933A
  - FL390

- **2017-03-04 19:38:49 UTC**: 10 minutes before CPA
  - IBK7VY
  - FL390
  - IBK6113
  - FL390
SAME FLOWS (FIG. 13)
CONCLUSION AND FUTURE WORKS

A robust method to produce a taxonomy of ATC situations

- Include altitude analysis, with BDS4.0 registers (MCP altitude setting), maybe even include TCAS RAs
- Take it to the production-level: construct a systematic catalogue of deconfliction situations
- Workload assessment: are there conservative/cautious and “wait-and-see” ATC behaviours?
- Network architecture simultaneously learning clusters and performing anomaly detection
The code is available at
https://traffic-viz.github.io/paper/sectflow.html

- Open source library providing data analysis methods commonly applied to trajectories and airspaces.
- Easy access to data from various sources (OpenSky Network, Eurocontrol DDR, etc.).
- Any help welcome! (bug reports, pull requests, etc.)