Combining Visual Analytics and Machine Learning for Route Choice Prediction
Application to Pre-Tactical Traffic Forecast

Rodrigo Marcos
Data Scientist, Nommon Solutions and Technologies

SESAR Innovation Days
Beograd, 29th November 2017
Scope and Objectives

Problem:
• ATFCM in the pre-tactical phase

Current approach:
• Based on similarity
  http://www.eurocontrol.int/articles/ddr-pre-tactical-traffic-forecast

Objectives:
• Use visual analytics to extract route choice determinants
• Model behaviour of airlines regarding route choice between airport pairs using machine learning techniques
• Evaluate pre-tactical prediction power
Abundant research on tactical trajectory prediction:
- Prediction of arrival time
- Conflict detection
- ...

Limited research on airline route choice prediction before the availability of flight plans (pre-tactical forecast):
Approach

• Data: actual trajectories (M3) from DDR2
• Route clustering per OD
• Visual exploration of route choice determinants
• Train a machine learning model
• Evaluate quality of predictions vs null model
Case Studies

• ODs:
  • Istanbul to Paris
  • Canary Islands to London

• Multinomial regression

• Candidate variables
  • Route length
  • Charges
  • Time
  • Schedule
  • Congestion

• Temporal scope:
  • Training/exploration: AIRACs 1601-1603
  • Testing: AIRACs 1501, 1502
Clustered with DBScan
Metric: Flown kilometres per ANSP
Visual Exploration Cost-worthiness

2 variables considered
- Average route length
- Average route charges

1 variable discarded
- Average flight time

<table>
<thead>
<tr>
<th>Cluster</th>
<th>No of flights</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>139</td>
</tr>
<tr>
<td>1</td>
<td>110</td>
</tr>
<tr>
<td>2</td>
<td>190</td>
</tr>
<tr>
<td>3</td>
<td>218</td>
</tr>
<tr>
<td>4</td>
<td>117</td>
</tr>
<tr>
<td>5</td>
<td>73</td>
</tr>
<tr>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
</tr>
</tbody>
</table>
Visual Exploration
Airline Behaviour

2 variables considered
• Arrival time
• Airline

20:00-22:00 (all airlines)
22:00-00:00 (all airlines)
Visual Exploration
Congestion

1 variable considered
- Average number of regulated flights

1 variable discarded
- Average standard deviation of en-route FL with respect to RFL
Visual Exploration
Cluster Properties

<table>
<thead>
<tr>
<th>Cluster</th>
<th>No of flights</th>
<th>Average length (NM)</th>
<th>Average charges (EUR)</th>
<th>Regulations per flight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>139</td>
<td>1277</td>
<td>1188</td>
<td>0.15</td>
</tr>
<tr>
<td>1</td>
<td>110</td>
<td>1314</td>
<td>1144</td>
<td>0.11</td>
</tr>
<tr>
<td>2</td>
<td>190</td>
<td>1273</td>
<td>1199</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>218</td>
<td>1274</td>
<td>1203</td>
<td>0.06</td>
</tr>
<tr>
<td>4</td>
<td>117</td>
<td>1256</td>
<td>1207</td>
<td>0.07</td>
</tr>
<tr>
<td>5</td>
<td>73</td>
<td>1274</td>
<td>1204</td>
<td>0.1</td>
</tr>
<tr>
<td>6</td>
<td>29</td>
<td>1271</td>
<td>1229</td>
<td>0.03</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>1304</td>
<td>1152</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Istanbul - Paris
Visual Exploration
Cluster Properties

Canary Islands - London

<table>
<thead>
<tr>
<th>Cluster</th>
<th>No of flights</th>
<th>Average length (NM)</th>
<th>Average charges (EUR)</th>
<th>Regulations per flight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>659</td>
<td>1620</td>
<td>1653</td>
<td>0.18</td>
</tr>
<tr>
<td>1</td>
<td>238</td>
<td>1638</td>
<td>1676</td>
<td>0.13</td>
</tr>
<tr>
<td>2</td>
<td>68</td>
<td>1740</td>
<td>1051</td>
<td>0.13</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>1732</td>
<td>1582</td>
<td>0.46</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>1724</td>
<td>1893</td>
<td>0.42</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>1780</td>
<td>1165</td>
<td>0</td>
</tr>
</tbody>
</table>
Approach
Parameters

Route parameters (used for modelling):

• Cost-worthiness:
  • Average route charges
  • Average route length

• Congestion:
  • Rate of regulated flights

Flight parameters (used for segmentation):

• Airline (CASK)
• Arrival time
Modelling Approach
Multinomial Regression Model

\[ P(Y_i = j) = \frac{e^{\beta_i \cdot X_j + \alpha_j}}{1 + \sum_{k=1}^{J} e^{\beta_i \cdot X_k + \alpha_k}} \]

Model of class i and cluster j
- \( X_j \) vector of parameters of cluster j
- \( \beta_i \) vector of constants of model i
- \( \alpha_j \) independent constant of cluster j

Variables:
- Cost-worthiness:
  - Average route charges
  - Average route length
- Congestion:
  - Rate of regulated flights
Approach

Training and Validation

Data

Clustering

Routes

Segmentation

Training 70%

Validation 30%

¿=?

Guess 2

Guess 0

Model 0

Model 1

Model 2

Model 3

Model 4

Model validation

SIDs, Beograd, 29th November 2017 – Combining Visual Analytics and Machine Learning for Route Choice Prediction
Validation Results
Canary Islands-London

- Low number of routes
- Very different
- Well explained
Validation Results
Istanbul-Paris

- High number of options
- Similar routes
- Missing explanatory variables?

<table>
<thead>
<tr>
<th>Cluster</th>
<th>No of flights</th>
<th>Average length (NM)</th>
<th>Average charges (EUR)</th>
<th>Regulations per flight</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>218</td>
<td>1274</td>
<td>1203</td>
<td>0.06</td>
</tr>
<tr>
<td>4</td>
<td>117</td>
<td>1256</td>
<td>1207</td>
<td>0.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster</th>
<th>No of flights</th>
<th>Average length (NM)</th>
<th>Average charges (EUR)</th>
<th>Regulations per flight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>139</td>
<td>1277</td>
<td>1188</td>
<td>0.15</td>
</tr>
<tr>
<td>5</td>
<td>73</td>
<td>1274</td>
<td>1204</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Approach

Testing

Dataset 2

Clustering → Routes

Dataset 1

Train → Model 0 → Model 1 → Model 2 → Model 3 → Model 4

Compare

Segmentation

Class 0 → Class 1 → Class 2 → Class 3 → Class 4

Route 0 → Route 1 → Route 2

SIDs, Beograd, 29th November 2017 – Combining Visual Analytics and Machine Learning for Route Choice Prediction
Testing Results
Canary Islands-London

• The model captures:
  • behaviour of new airline (Norwegian)
  • airlines changing route options
  • Improvements with respect to null model
Testing Results
Istanbul-Paris

- The model captures:
  - other routes considered (7)
  - significant change in charges
  - Much better than null model

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Charges (train)</th>
<th>Charges (testing)</th>
<th>Regulations (train)</th>
<th>Regulations (testing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1188</td>
<td>1305</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>1204</td>
<td>1260</td>
<td>0.07</td>
<td>0.02</td>
</tr>
</tbody>
</table>

SIDs, Beograd, 29th November 2017 – Combining Visual Analytics and Machine Learning for Route Choice Prediction
Applicability

- Potential for pre-tactical demand forecast
- Range of applicability needs to be clearly identified:
  - Training data requirements
  - Prediction error measurement
  - Generalisation to other ODs
Future Research Directions

- Better explanatory variables
  - Other indicators
  - Congestion as a function of time
  - Other flight inputs: wind, type of regulation, route availability...
- Training with several years’ data
- Continuous training/prediction (automatic adaptive training data)
- Combination with model-based approaches (cost optimisation)
SIDs, Beograd, 29th November 2017
Combining Visual Analytics and Machine Learning for Route Choice Prediction
Application to Pre-Tactical Traffic Forecast

Thank you very much for your attention!