Traffic Flow Management
Data Mining Update

Shon Grabbe
TFM Technical Lead
Research Objective

Given relevant NAS status, weather and traffic information determine the “best” collection of Traffic Management Initiatives to implement for day-of-operation planning purposes.

Work Sponsored By:
Richard Jehlen
Director, Planning and Procedures
FAA System Operations
Approach

**Historical NAS Status, Weather and Traffic Information**

**Airport and Airspace Weather Translation**

**Similar Days in the NAS**

**Historical TMIs**

**Parse/Aggregate/Digitize TMIs**

**Similar TMIs used under similar days in the NAS**
Approach

Historical NAS Status, Weather and Traffic Information

Airport and Airspace Weather Translation

Similar Days in the NAS

Historical TMIs

Parse/Aggregate/Digitize TMIs

Similar TMIs used under similar days in the NAS
Convective weather observations and forecasts are obtained through Corridor Integrated Weather Service (CIWS) archives.

Airport weather observations and forecasts obtained from METAR and TAF reports.

Aircraft Situation Display to Industry (ASDI) currently being used for traffic data.
Approach

Historical NAS Status, Weather and Traffic Information

Airport and Airspace Weather Translation

Similar Days in the NAS

Historical TMIs

Parse/Aggregate/Digitize TMIs

Similar TMIs used under similar days in the NAS
Simplified version of the Weather Impacted Traffic Index (WITI) used to estimate the impact of convective weather on en route air traffic.

\[ WITI(k) = \sum_{1 \leq j \leq m} \sum_{1 \leq i \leq n} T_{i,j}(k)W_{i,j}(k) \]

- \( W_{i,j} = 1 \) at (i, j) locations with severe weather
- \( T_{i,j} = \) Traffic counts at (i, j) location
Color indicates the number of times that a particular reroute was used when the WITI in a particular Center was above a specified threshold value.

Volume related reroutes.

Extensive routing on the east coast.

- Daily WITI used to capture the en route, convective weather impacts
- Additional features/attributes required for non-weather related reroutes (e.g., airport queuing delays)
Approach

Historical NAS Status, Weather and Traffic Information

Airport and Airspace Weather Translation

Similar Days in the NAS

Similar TMIs used under similar days in the NAS

Historical TMIs

Parse/Aggregate/Digitize TMIs
Clustering techniques can be used to reduce the number of en route features that are considered and subsequently simplify the learning automation algorithms.

- Daily WITI of 20 Centers and Total NAS WITI used
- 321 days in 2011 used as input data
- WITI distribution obtained
  - Mean, variance, quartiles, 90th, and 95th percentiles
- WITI values standardized
- Perform cluster analysis. Extracted 17 clusters
- Attribute each day to a specific cluster
  - Total NAS WITI used to qualify a day as mild/moderate/severe weather
  - Percentile values of Center WITI used to classify regional weather
Cluster 13: Severe Weather Day

Number of Matching Days: 5
Cluster 12: Severe Weather Day

Number of Matching Days: 4
Cluster 11: Moderate Weather Day

Number of Matching Days: 4
Weather Cluster to Reroute Mapping

Cluster 13: Severe Weather Day
Cluster 12: Severe Weather Day
Cluster 11: Moderate Weather Day
Approach

1. **Historical NAS Status, Weather and Traffic Information**
2. **Airport and Airspace Weather Translation**
3. **Similar Days in the NAS**

- **Historical TMIs**
  - Parse/Aggregate/Digitize TMIs
  - Similar TMIs used under similar days in the NAS
NTML and the fly.faa.gov website are the primary sources of historical reroutes, MIT restrictions, AFPs, Ground Delay Programs and Ground Stops

- Much of the data is not directly machine readable

- This effort will ultimately be a client for the “TMI Cube” but currently much of the work is occurring in parallel
Approach

Historical NAS Status, Weather and Traffic Information

Airport and Airspace Weather Translation

Similar Days in the NAS

Historical TMIs

Parse/Aggregate/Digitize TMIs

Network Representation of Reroutes

Similar TMIs used under similar days in the NAS
• 989 “unique” ATCSCC ADVZY reroutes issued in 2011
• Data reduction techniques required to reduce this number for describing the high-level, national-plan
One technique to reduce the number of reroutes is to represent them in terms of a collection of pre-published playbook routes.
There are over 200 playbook routes, so clustering the routes would improve computation performance.

One approach to clustering routes entails representing each route in terms of the Centers which they pass through.

List of all possible NAS network edges: ZSE-ZOA, ZOA-ZLA, ZSE-ZLC, ZOA-ZLC, etc.

Network edges used by “BUM”: ZSE-ZLC, ZOA-ZLC, ZLC-ZDV, ZLA-ZDV, ZLA-ZAB, etc.

Vector representing the “BUM” network edges: <0, 0, 1, 1, 0, 1,…>
Candidate attributes of the playbook networks

<table>
<thead>
<tr>
<th>ZSE-ZOA</th>
<th>ZOA-ZLA</th>
<th>ZSE-ZLC</th>
<th>ZOA-ZLC</th>
<th>ZLA-ZLC</th>
<th>...</th>
<th>CZM-ZBW</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>...</td>
<td>0</td>
</tr>
</tbody>
</table>

- K-means algorithm used to cluster the playbook routes
- Number of clusters determined by maximizing the average silhouette score of all clusters
- Clustering reduced the number of routes from 200+ to 17 playbook clusters
Sample members of “Cluster 3” - 7 members in cluster

Route nodes and edges for CAN_1_EAST
Threshold = 0

Route nodes and edges for CAN_2_EAST
Threshold = 0

Route nodes and edges for CAN_3_EAST
Threshold = 0

Route nodes and edges for CAN_5_EAST
Threshold = 0

CAN East Routes
Sample members of “Cluster 4” - 4 members in cluster

Route nodes and edges for MEM_GQE
Threshold = 0

Route nodes and edges for DC_NORTH
Threshold = 0

Route nodes and edges for ROCKIES_1
Threshold = 0

Route nodes and edges for ROCKIES_4
Threshold = 0

East to West Routes
Sample members of “Cluster 6” - 18 members in cluster

West to East Routes
Approach

Historical NAS Status, Weather and Traffic Information

Airport and Airspace Weather Translation

Similar Days in the NAS

Historical TMIs

Parse/Aggregate/ Digitize TMIs

Grid-based Representation of Reroutes

Similar TMIs used under similar days in the NAS
Historical reroutes mapped onto a uniform grid to facilitate clustering, as part of the “Regional Plan”
Sample Grid-based Routes

Sample Reroute
Start at circle end at square

Route nodes compared when clustering
53 un-clustered routes, start at circle end at square

Routes clustered using a k-means clustering algorithm
Sample Reroute Clusters

Cluster 1

Cluster 2
Sample Reroute Clusters

Cluster 3

Cluster 4
Approach

Historical NAS Status, Weather and Traffic Information

Airport and Airspace Weather Translation

Similar Days in the NAS

Historical TMIs

Parse/Aggregate/Digitize TMIs

Similar TMIs used under similar days in the NAS
• **Objective:** Develop decision association models based on historical data using data mining and other learning automation concepts
  – Given a weather disruption signature, what national plans consisting of GDPs were used in the past to deal with similar event?
• Data (NTML, ASPM, METAR from 2008-2010)
• Decision tree learning algorithms
GDP Locations

- GDPs at major airports (EWR, LGA, JFK, PHL, BOS, SFO, ATL, ORD):
  - Having GDPs at all 8 airports on the same day is rare,
  - 70% of the days have GDPs at more than one of these airports
- Developed two types of models
  - Models predicting GDPs at individual airports
  - Models characterizing interdependence between GDPs at different airports
• Feature Selection: Best discriminating feature will produce sets with high probability of a control action.
• Each path from the root of the tree to a leaf represents a weather signature.
Capability 1: Predict GDPs at individual airports under specified conditions

For weather signature on 12/11/2010, what is the likely GDP plan based on the historical pattern of GDPs?

<table>
<thead>
<tr>
<th>Airport</th>
<th>Predicted Probability of GDP</th>
<th>Actual GDP status</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWR</td>
<td>18%</td>
<td>No</td>
</tr>
<tr>
<td>LGA</td>
<td>14%</td>
<td>No</td>
</tr>
<tr>
<td>JFK</td>
<td>14%</td>
<td>No</td>
</tr>
<tr>
<td>PHL</td>
<td>43%</td>
<td>No</td>
</tr>
<tr>
<td>BOS</td>
<td>3%</td>
<td>No</td>
</tr>
<tr>
<td>ORD</td>
<td>53%</td>
<td>Yes</td>
</tr>
<tr>
<td>ATL</td>
<td>6%</td>
<td>No</td>
</tr>
<tr>
<td>SFO</td>
<td>86%</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Capability 2: Capture interdependence between GDPs at different airports

- Predict number of GDPs in New York region
  - Likely number of GDPs in New York area is 3 on 12/12/2010 with probability of 65%

- Predict GDP patterns in New York region
  - Likely regional plan predicted with weather on 12/12/2010
    - 65% probability of GDPs at JFK, EWR, LGA
    - 9% probability of GDPs at JFK and EWR
    - 8% probability of GDPs at LGA and EWR
  - Likely regional plan predicted with weather on 12/15/2010
    - 32% probability of GDPs at JFK, EWR, LGA
    - 26% probability of GDPs at JFK and EWR
    - 22% probability of GDPs at LGA and EWR
    - 12% probability of GDP at EWR

- Predict Probability of a GDP at another airport given that there is a GDP at an airport
  - Northeast airport GDPs correlated with each other. This is partly because weather at these airports is correlated with each other.
  - Example: Conditional probabilities of GDP at other airports given that there is GDP at EWR are as follows:
    - LGA: 0.55  JFK: 0.56  PHL: 0.37  BOS: 0.32
  - In some airport pairs, there are temporal patterns regarding which airport has GDP first. For example, on days with both LGA and EWR GDPs, EWR first 45 days, LGA first 149 days
• Understanding variation in control actions under similar weather conditions
  – Possible causes of variation may include controller subjectivity, differing goals, availability of alternate control actions / cancellations, missing information about impacting factors
  – Therefore, a study comparing official GDP causes with historical patterns in decision trees on days with following weather signature was done.
    • Visibility < 3.5 nm, wind speed < 11, number of GDPs in rest of North East USA major airports < 3
    • Likely Control Action on these days is GDP
## Days with Low visibility

<table>
<thead>
<tr>
<th>Date</th>
<th>MCWiti</th>
<th>Highwind witi</th>
<th>Vis</th>
<th>Ceil</th>
<th>Scharr</th>
<th>MC</th>
<th>Wsp</th>
<th>North East GDP</th>
<th>Actual GDP</th>
<th>GDP cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/4/08</td>
<td>39</td>
<td>0</td>
<td>1</td>
<td>700</td>
<td>24</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>Low vis</td>
</tr>
<tr>
<td>2/6/08</td>
<td>11</td>
<td>8</td>
<td>3</td>
<td>1600</td>
<td>23</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>Wind</td>
</tr>
<tr>
<td>2/17/08</td>
<td>23</td>
<td>0</td>
<td>2</td>
<td>600</td>
<td>22</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2/23/08</td>
<td>30</td>
<td>0</td>
<td>2</td>
<td>700</td>
<td>20</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>Low Ceiling</td>
</tr>
<tr>
<td>2/29/08</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>500</td>
<td>25</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4/5/08</td>
<td>9</td>
<td>8</td>
<td>2</td>
<td>2200</td>
<td>20</td>
<td>2</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4/9/08</td>
<td>24</td>
<td>0</td>
<td>1</td>
<td>300</td>
<td>25</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>Low vis</td>
</tr>
<tr>
<td>4/10/08</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>200</td>
<td>26</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4/12/08</td>
<td>10</td>
<td>24</td>
<td>0</td>
<td>200</td>
<td>20</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>Wind</td>
</tr>
<tr>
<td>5/3/08</td>
<td>41</td>
<td>0</td>
<td>0</td>
<td>300</td>
<td>20</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>Low vis</td>
</tr>
</tbody>
</table>
• Probabilistic GDP predictions are accurate within 10% accuracy.
• Actual cause is attributed to a single weather factor such as wind, ceiling or visibility whereas some of the days have multiple adverse factors. Decision tree signature patterns often include more than one factor.
• We do not have data on some of the GDP causes such as ‘runway construction and maintenance’
• In some cases, data we have does not explain stated GDP causes. For example, 2/6/2008 seem to have low visibility conditions and low wind speed conditions at EWR. While existence of GDP on that day is consistent with historical pattern, the attribution of GDP cause to wind is puzzling.