

NASA Data Mining Update

Jessica Nowinski

Human Factors REDAC Meeting March 29, 2016

NASA Data Mining Efforts



- Anomaly Detection
 - Anomaly discovery over large set of variables
 - Given all flights within a set of common parameters (location, phase of flight) which flights are statistically most different from the others
 - Some false positives
- Precursor Identification
 - Given undesirable effect (e.g., go-around), identify
 precursors (e.g., overtake situation, high speed approach)
- Text mining
 - Text classification, topic identification



Exceedance-Based Methods

- Known anomalies
- Conditions over 2-3 variables (e.g., speed > 250 knots, altitude = 1000 ft, landing)
- Cannot identify unknown anomalies
- Low false positive rate, high false negative (missed detection) rate.

Data-Driven Methods



- Discover anomalies by
 - learning statistical properties of the data
 - finding which data points do not fit (e.g., far away, low probability)
 - no background knowledge on anomalies needed
- Complementary to existing methods
 - Low false negative (missed detection) rate
 - Higher false positive rate (identified points/flights unusual, but not always operationally significant)
- Data-driven methods -> insights -> modification of exceedance detection

Application of MKAD Anomaly Detection Too

- Collaboration with airline partners
 - Developed algorithm to incorporate discreet and continuous flight data, text data and weather
 - Considered particular variables of interest, such as fuel burn
 - Compare model of fuel burn with actual instantaneous fuel burn
 - Where difference is high, problem may be occurring
- Extension to analysis of NAS data using the same tool
 - Started with LAX
 - Expanded to Denver (complexity due to weather), NYC Metroplex (complexity due to traffic) and Houston (metroplex redesign)

Anomaly: JFK High Speed Go-Around



- Overshoots Extended Runway Centerline (ERC) by over 1 SM
- Over 250 Kts @2500 Ft.
- Angle of intercept > 40°
- Overshoots 2nd approach



Anomaly: JFK Recurring High Vertical Speed Descent



- Same small B190 cargo aircraft over 7 different days to JFK.
- Unconventional maneuvers to arrest airspeed and altitude.
- Tool could help identify non SOP flights to help controllers amend procedures so that all controllers are familiar with these types of approaches.



Anomaly: DEN Aggressive Runway Change



- Potential overtake averted
- Resulted in unstable approach on cross runway.
- "Belly-Up" situation with FLT899 on turn to final inside the outer marker.



Anomaly: NE Landing East Flow Aug 28 (after metroplex redesign)





Precursor to go-around--Energy Mismanagement





Upper left plot shows that ground speed is higher than normal. This is responsible for precursor index (lower right) being high, indicating high probability of go-around. Algorithm automatically identified precursor, given goaround condition as target effect.

Ongoing Implementation Work



- Implement anomaly detection on PDARS. Prototype daily reporting system
- Implement within Sherlock Data Warehouse for daily and hourly reporting
- Set up MKAD within Aviation Safety Information Analysis and Sharing (ASIAS) sandbox for anomaly detection demonstration
- Broaden from 1-2 flights to airspace metrics
- Semi-automatically identify effects of changes like Houston Metroplex