Reasoning and Learning in Real-World Systems for Long-Term Autonomy (LTA)
Papers from the AAAI 2018 Fall Symposium

Big Data and Deep Learning Models for Automatic Dependent Surveillance Broadcast (ADS-B)

Ying Zhao
Naval Postgraduate School
yzhao@nps.edu

Richard Wu
UMass Dartmouth
rwu@umassd.edu

Andrew Polk
UC Santa Barbara
polk@umail.ucsb.edu

Matthew Xi
Indiana University Bloomington
mxi@iu.edu

Tony Kendall
Naval Postgraduate School
wakendal@nps.edu

Abstract

ADS-B functions with satellite (GPS) rather than radar technology to more accurately observe and track air traffic. Aircraft equipped with an ADS-B Out transmitter sends position, altitude, heading, ground speed, vertical speed, call sign, and other aircraft information to a network of ground stations that relays the information to air traffic controllers and other aircraft. Our work in progress applies various big data tools and deep learning models such as convolutional neural networks to process and use the ADS-B big data to predict if a flight is commercial or military.

Introduction

ADS-B functions with satellite (GPS) rather than radar technology to more accurately observe and track air traffic. Aircraft equipped with an ADS-B Out transmitter sends their position, altitude, heading, ground speed, vertical speed, call sign, and other aircraft information to ground stations that relays the information to air traffic controllers and those who have ground ADS-B receivers. Pilots of aircraft equipped with a receiver for optional ADS-B receive traffic and meteorological information. Aircraft operating in most controlled U.S. airspace must be equipped for ADS-B Out by January 1, 2020 (FAA 2018).

With ADS-B operational across the country, pilots in equipped aircraft have access to air traffic services that provide a new level of safety, better situational awareness, and more efficient search and rescue.

Data Description

An aircraft can be identified by radar transponder identification, friend or foe (IFF) modes such as one, three, and five (military). This is done by Line of Sight (LOS) air traffic control ground radar stations. For improved (cooperative) surveillance for flight separation and control an aircraft can have an Automatic Dependent Surveillance-Broadcast (ADS-B) Out system to broadcast its identification and location from the aircrafts GPS to LOS receiving ground stations and other aircraft (around a 150 mile range). According to International Civil Aviation Organization (ICAO), a.k.a, the international "FAA", states that notable outcomes of using ADS-B include a new frequency allocation for space-based ADS-B reception, enabling tracking of aircraft globally including remote and polar regions. Plans are to fully employ ADS-B and the related infrastructure by 2020 (FAA 2018).

Table 1: 3-month ADS-B Data Divided into 7 Data Sets

Name	Month of Data
Data set 7_1	July, 2016,Train
Data set 7_0	July, 2016,Test
Data set 7_2	July, 2016,Test
Data set 8_0	August, 2016,Test
Data set 8_1	August, 2016,Test
Data set 6_0	June, 2016,Test
Data set 6_1	June, 2016,Test

Four terabytes ADS-B data (ADSBexchange.com 2017), were downloaded from the website ADS-B Exchange sampled every minute, for the whole year (6/2016 to 6/2017) and processed the selected fields and geo-spatial areas for a three month time period using the NPS high performance computer. The data are divided into 7 data sets representing about 10 days of the flights for a selected region as shown Table 1. There are approximately 150,000 flights (tracks) in each data set. Our goal is to build ML/AI models such as lexical link analysis (LLA) and reinforcement learning (RL) algorithms to rapidly and accurately classify military and commercial aircraft based on kinematic characteristics. The objective of this paper is to investigate how the track data could be represented as images and fed into the deep learning algorithm such as convolutional neural networks (CNN) for high classification accuracy.

There are two versions of all the three data sets

- Data: Original kinematic attributes with numeric values for a whole track as follows:
  - Mean altitude (barometric)
  - Mean altitude change
  - Mean altitude change absolute
  - Mean speed
  - Mean speed change
  - Mean speed change absolute
  - Mean heading change
  - Mean heading change absolute
  - Total altitude change

– Total altitude change absolute
– Total heading change
– Total heading change absolute
– Total speed change
– Total speed change absolute
– Total duration

The target attribute to predict is mil_true: if an airborne object is military (true) or commercial (false). The percentage of military flights is about 2%.

- Unique data: Discretized original kinematic at-tributes. For example, the “average altitude attribute is turned into the following categorical attributes:
  – Likelihood to be commercial when average altitude between 16,607 feet and 29,525 feet
  – Likelihood to be military when average altitude between 16,607 feet and 29,525 feet
  – Likelihood to be commercial when average altitude between 3,689 feet and 16,607 feet
  – Likelihood to be military when average altitude between 3,689 feet and 16,607 feet
  – Likelihood to be commercial when average altitude less than 3,689 feet
  – Likelihood to be military when average altitude less than 3,689 feet
  – Likelihood to be commercial when average altitude more than 29,525 feet
  – Likelihood to be military when average altitude more than 29,525 feet

  The values of discretization are computed automatically by examining the means and standard deviations of the original attributes. After discretization, the percentage of military flights is about 6%.

Methods

Data Exploration

Tableau (tab) can filter the data with any of the metrics included in the data set. Changing the sizes of the nodes based on the number of virtual tracks that pass through that point allows for the visualization of heavy traffic paths over a map. The example below clearly shows the most popular flight paths based on the darker lines on the map. Tableau also allows for coloring the points on the map though other parameters in the data. Tableau can display virtual tracks and statistics. Figure 1 shows examples of virtual tracks of aircraft for a selected region when the altitudes are higher than 18,000 feet for the training set 7_1. Figure 2 shows the virtual tracks when the tracks altitudes are less than 18,000 feet and total heading change for tracks are larger than 2000. Heading is the compass direction toward which an airborne object should be moving. The blue tracks are commercial and the orange tracks are military. Comparing Figure 1 and Figure 2, we can see that the kinematic characteristics for Figure 2 (i.e., flying below 18,000 feet and total heading change greater than 2000) are more likely to indicate military aircraft, while the kinematic characteristic (i.e., flying higher than 18,000 feet) for Figure 1 is more likely to indicate commercial aircraft.

Supervised Learning Using Weka

We also reviewed and tested supervised learning methods such as decision trees, logistic regression, nave Bayes, nearest neighbors, and random forest models in the open source data mining and machine learning tool Weka (Hall et al. 2009) as shown in Figure 3. K-nearest neighbors and random forest perform the best for both versions of the data. The part of the research gave us the baselines for comparison for the data set. Since ADS-B data are mostly commercial flights, even 2% flights are self-reported as military, these aircraft mostly fly similar to commercial aircraft based on their kinematic characteristics. This is the key reason for the difficulty of the classification task of this data set.

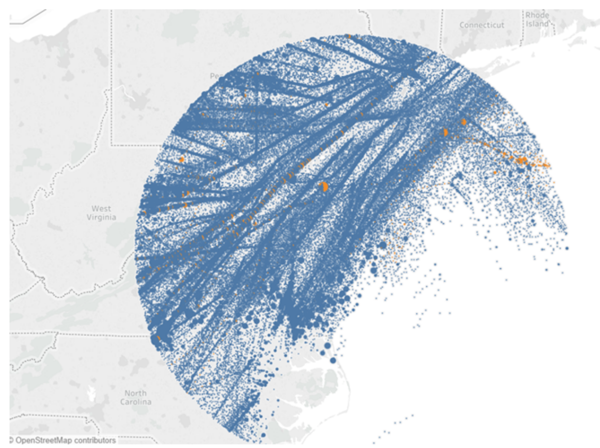


Figure 1: Sample virtual commercial tracks with altitude greater 18,000 feet

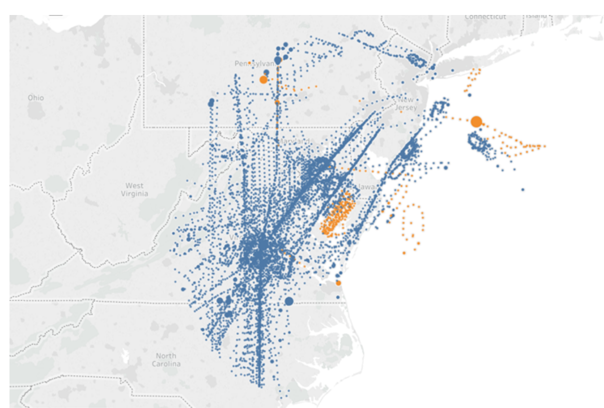


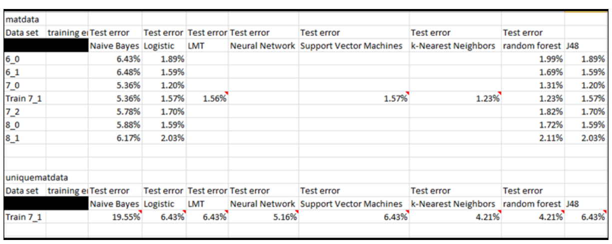
Figure 2: Map with altitude below 18,000 feet and total heading change greater than 2000. The blue tracks are commercial and the orange tracks are military

Convolutional Neural Networks (CNN)

CNN can train themselves to classify objects in images with high accuracy (CNN 2018). We explored if CNN can be used to predict military and commercial flights from the ADS-B track data. We represented the ADS-B track data as images and then apply CNN. The ADS-B data set contains a flights information such as the aircraft name, timestamp, speed, heading, altitude, military or commercial. We define a track to be the period between takeoff and landing. For each track, we focus on three in-put attributes i.e. speed, heading, and altitude and one output attribute i.e. its classification as military or commercial. Each track updates its kinematic data at one-minute time intervals.

For this particular data set, the color scale will represent altitude values ranging between -16,935 and 126,400. The negative altitudes are probably data errors. Figure 5 shows an example of a track heat map using the color scale in Figure 4 and altitude is the color or heat.

Segmenting Data into Tracks Microsoft Excel was used to open the data set as a .csv file. A python script was created to split the ADS-B data into tracks, or flight paths, and two data folders. One folder for all the civilian aircraft and one for military. The complete data was presorted by time and by aircraft ID. The python script split the full data set into smaller ones based on aircraft ID and then again based on the time between measurements. If the time between sensor measurements for the aircraft was greater than one hour a new track was formed. Afterward, another python program is written to plot each flight into one figure. The main problem with the flight heading data was the heading value was absolute and not relative causing the data to jump when



matdata									
Data set	training e	Test error	Test error	Test error	Test error	Test error	Test error	Test error	
[redacted]		Naive Bayes	Logistic	LMT	Neural Network	Support Vector Machines	k-Nearest Neighbors	random forest	J48
6_0		6.43%	1.89%					1.99%	1.89%
6_1		6.48%	1.59%					1.69%	1.59%
7_0		5.36%	1.20%					1.31%	1.20%
Train 7_1		5.36%	1.57%	1.56%		1.57%	1.23%	1.23%	1.57%
7_2		5.78%	1.70%					1.82%	1.70%
8_0		5.88%	1.59%					1.72%	1.59%
8_1		6.17%	2.03%					2.11%	2.03%
uniquematdata									
Data set	training e	Test error	Test error	Test error	Test error	Test error	Test error	Test error	
[redacted]		Naive Bayes	Logistic	LMT	Neural Network	Support Vector Machines	k-Nearest Neighbors	random forest	J48
Train 7_1		19.55%	6.43%	6.43%	5.16%	6.43%	4.21%	4.21%	6.43%

Figure 3: Supervised learning methods and error rates used for baselines and comparison

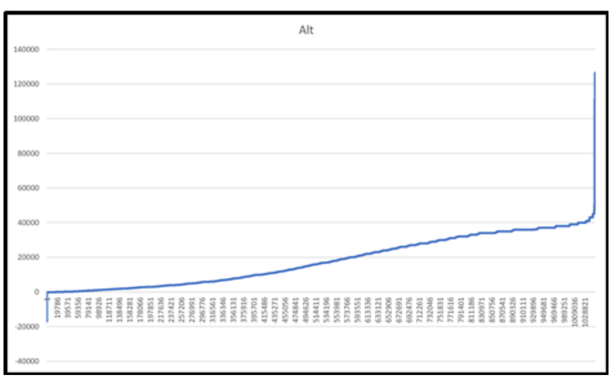


Figure 4: A graph showing the minimum and maximum altitude range over 10 days of ADS-B data from July 10th, 2016 to July 20th, 2016

Figure 5: Sequential Color scale representing low altitudes on the left (purple) and high altitudes on the right (yellow)

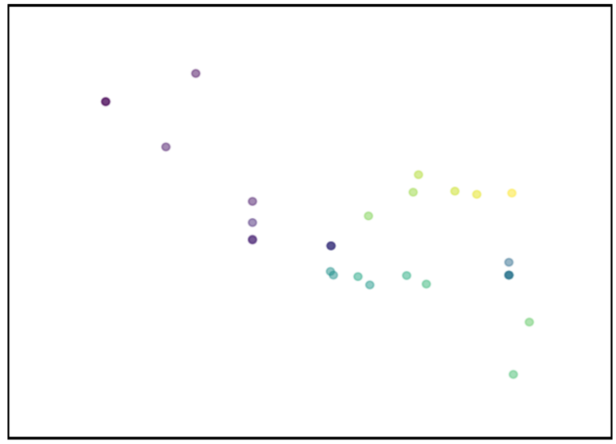


Figure 6: Heat map of speed (x-axis), heading (y-axis) and altitude (color) for one track

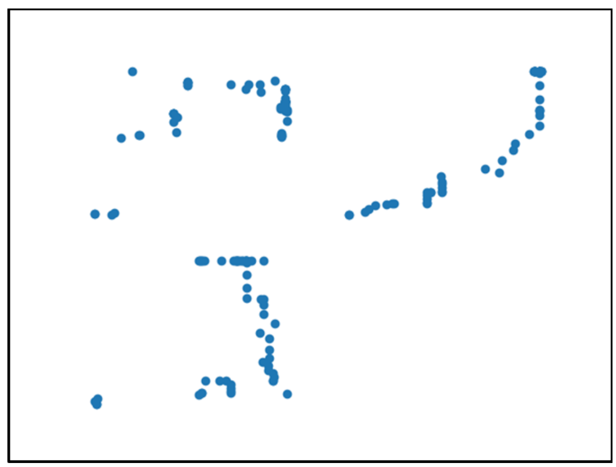


Figure 7: Three subplots demonstrating three different relations between speed, heading, and altitude

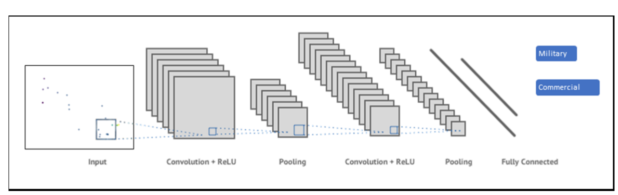


Figure 8: The CNN Architecture demonstrating a scatter plot to be classified as military or commercial

crossing the 0°or 180°. To smooth the track headings in the data was modified to fit between 90°and 90°. A full rotation of the heading represented this way is visually equivalent to that of a sine wave. Some data is lost in this representation as there is no way to discern 0 °from 180 but the relative change in heading between data points is saved. This smoothing method solves the problem of data jumping around when the heading changes from 1°to 359°as this will be shown as a change of 2°rather than 358°allowing for a cleaner visualization of the data for processing.

Transform Tracks into Images Before using a CNN, first we converted a whole track into an image. We represent the input attributes for a track as an image by plotting their relationship i.e. speed vs heading vs altitude as a color scale as shown in Figure 5 The color bar and axis of the graph have been turned off to reduce unnecessary noise in the image. A perceptually uniform sequential color map (Colormaps 2018) has been chosen to represent the altitude data. Sequential color maps change in lightness and saturation incrementally, using a single hue, and is suitable for information that has ordering as shown in Figure 4.

Figure 4 is a graph showing the minimum and maximum altitude for the train data set which is used to the color scale in Figure 4. The color scale represents the altitudes within the entire ADS-B data set, not within a single flight path. This is so that when comparing two flight paths (i.e. two plotted images) the color at a specific altitude would be represented as the same color.

In addition to the heat plot, we also tried to represent three pairwise (speed vs heading, heading vs altitude, altitude vs speed) kinematic feature scatter plots into one figure as another way to transform track data into images as shown in Figure 7. The top right section displays speed vs heading. The top left displays speed vs altitude, and the bottom left displays heading vs altitude. The borders of the subplot have been turned off to reduce the noise in the image.

Running CNN

For both methods, a whole track as a graph is re-sized from a 640x480 image to an 80x60 image. The sample ADS-B data for a testing CNN run contains 50,293 images of commercial flights and 2,316 images or 4.4% of military flights. The images were prepared to be fed into the CNN to classify commercial and military flights as shown in Figure 8.

The CNN in TensorFlow (Keras) performed ten iterations of training on both transformed images to produce an accuracy of 95.6% for classifying commercial flights.

We explored various big data and deep learning methods to classify ADS-B data sets. Supervised learning methods show certain ML accuracy. We also explored how to represent track data as images and then applied CNN for classifications. Our work in progress did provide promise that various big data tools and deep learning models such as convolutional neural networks can process and use the ADS-B big data to predict if a flight is commercial or military. Other software tools such as Tableau, Weka and even MS Excel show the potential that ADS-B data could be manipulated to identify commercial and military behavior to predict type of aircraft.

Conclusion and Future Work

The result of CNN remains a work in progress. Future work is needed, for example, using more data from different sources to enrich military aircraft. Also while our tools demonstrated some promising results we need to explore tools that handle big data and near real time data such as BDP (Big Data Platform) which is being used in related work.

Acknowledgements

Authors would like to thank the Naval Research Program at the Naval Postgraduate School and the Naval Research Enterprise Internship Program at the Office of Naval Re-search for the research support. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied of the U.S. Government.

References

ADSBexchange.com, L. 2017. Ads-b exchange.

CNN. 2018. Convolutional neural network, retrieved from http://cs231n.github.io/convolutional-networks/.

Colormaps. 2018. Colormaps,retrieved from https://matplotlib.org/tutorials/colors/colormaps.htmlsequential.

FAA. 2018. FAA,retrieved from https://www.faa.gov/nextgen/how_nextgen_works/new_technology/.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; and Witten, I. H. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations* 11(1):10–18.

Tableau, https://www.tableau.com/.