DEVELOPMENT OF INCREASINGLY AUTONOMOUS TRAFFIC DATA MANAGER USING PILOT RELEVANCY AND RANKING DATA

Lisa R. Le Vie
Vincent E. Houston
NASA Langley Research Center
Hampton, VA

NASA's Safe Autonomous Systems Operations (SASO) project goal is to define and safely enable all future airspace operations by justifiable and optimal autonomy for advanced air, ground, and connected capabilities. This work showcases how Increasingly Autonomous Systems (IAS) could create operational transformations beneficial to the enhancement of civil aviation safety and efficiency. One such IAS under development is the Traffic Data Manager (TDM). This concept is a prototype 'intelligent party-line' system that would declutter and parse out non-relevant air traffic, displaying only relevant air traffic to the aircrew in a digital data communications (Data Comm) environment. As an initial step, over 22,000 data points were gathered from 31 Airline Transport Pilots to train the machine learning algorithms designed to mimic human experts and expertise. The test collection used an analog of the Navigation Display. Pilots were asked to rate the relevancy of the displayed traffic using an interactive tablet application. Pilots were also asked to rank the order of importance of the information given, to better weight the variables within the algorithm. They were also asked if the information given was enough data, and more importantly the "right" data to best inform the algorithm. The paper will describe the findings and their impact to the further development of the algorithm for TDM and, in general, address the issue of how can we train supervised machine learning algorithms, critical to increasingly autonomous systems, with the knowledge and expertise of expert human pilots.

Air traffic within the National Airspace System (NAS) is ever-increasing and "although humans today remain more capable than machines for many tasks, natural human capacities are becoming increasingly mismatched to the enormous data volumes, processing capabilities, and decision speeds that technologies offer or demand" (United States Air Force, 2010) Recognizing these challenges, NASA's Safe Autonomous Systems Operations (SASO) project's objectives are focused on developing technologies that enhance the safety and efficiency of civil aviation. Increasingly Autonomous Systems (IAS) are one avenue that could prove vital in decreasing a crew's workload, while enhancing safety and efficiency during the NextGen and other possible future airspace environments.

Utilizing IAS within the cockpit begins with understanding what an increasingly autonomous system is and what technologies are needed to profoundly improve a flight crew's overall awareness while maintaining or even decreasing workload. Autonomy allows an agent, human or machine, to act independently within a circumscribed set of goals; delegating responsibility to the agent(s) to achieve the overall system objective(s). (National Research Council, 2014). IAS lie within the sophisticated progression of current automated systems toward full autonomy. These systems, working together with humans, are expected to improve the safety, reliability, costs and operational efficiency of civil aviation. Implementation of IAS is imminent, which makes the development and proper performance of such technologies vital. The challenge is to develop these human-autonomy teams/systems where the combination of machine learning and human expertise exceeds the performance of either system alone.

To that end, an effort to develop cutting edge technology addressing an emerging airspace need as well as to serve as an IAS testbed for development and evaluation was created. The Traffic Data Manager (TDM) is an application that parses and displays traffic of interest while eliminating the clutter of insignificant surrounding traffic data. The application arises from an optimized Data Comm environment end-state where operations will become "net-centric" - as the transmittance of command, control, state, and intent information is passed autonomously between computers (agents) for efficient operational coordination and execution. Voice communication, between humans, will become non-existent as they can become a bottleneck to capacities. Nonetheless, to provide requisite human oversight, awareness, and intervention, an IAS is needed to effectively and concisely inform humans of "ownship-relevant" information (traffic, intent, messaging) being passed within this net-centric environment. The TDM application becomes an "intelligent party-line" process, only presenting (visual, aural, etc.) the information that the human must know to maintain the requisite awareness for possible subsequent action or intervention.

This technology relies on Machine Learning (ML) algorithms to parse all nearby traffic data, displaying only relevant data to the pilot. The primary component of the system is the TDM algorithm. TDM currently uses a supervised learning algorithm that relies on an Ensemble Learning framework (DeCoursey, 2003; Hastie, Tibshirani, & Friedman, 2009) where there are several methods for blending results into a very high-quality ensemble predictor. The fundamental challenge of IAS design and of this TDM application, in particular, is how to capture human expertise and knowledge and then effectively implement this knowledge within a machine learning architecture.

Data Collection Effort

Essential to the development of the machine learning algorithms was the collection of data needed to train the algorithm. A Dynamic Air Traffic Application (D.A.T.A) was developed and integrated into an EFB-like framework. Real-time flight data was randomly assigned a latitude and longitude and placed within a range of 20 or 40 nautical miles from the ownship. These data points were then displayed to the pilot in groups of 20, as shown in Figure 1a, and they were asked to rate the relevancy of the selected aircraft in relation to their own. When an aircraft was selected, a box appeared in the lower left hand corner (enlarged to enable viewing in Figure 1b) giving the selected aircraft's identification, type, altitude, speed, heading and vertical trend.

a. b.

Figure 1. D.A.T.A. Screenshots

Thirty-one pilots, all current or recently retired with an Airline Transport Pilot rating, were asked to choose a relevancy (relevant, maybe relevant, or not relevant) for each selected aircraft. This was repeated for each aircraft and each scenario. Each pilot saw 36 scenarios with 20 aircraft per scenario. Over 22,000 data points, with their selected relevancies, were collected from the pilots to be used in training the TDM algorithm.

Training TDM

The TDM supervised learning algorithm was initially trained using 75% of the 22,000 data points. These data points consist of aircraft state data that the pilots considered important to determining relevancy (i.e., course, heading, airspeed, altitude, range, bearing, etc.) as well as the pilot reported relevancy of the aircraft to the ownship's position. Testing was done with the remaining 25%, with the pilot-reported relevancy removed. The relevancy determination is only necessary to TDM in its training phase. Algorithm training took place using a MatLab Machine Learning, Tree-Bagger ensemble. The algorithm utilizes an embedded supervised learning algorithm to eliminate insignificant surrounding traffic, highlight traffic of interest or note, and identify operational significance autonomously.

Further Considerations

At the end of data collection, pilots were asked which of the two, heading or range, being the two salient parameters of the Navigation Display, was their primary consideration when choosing an aircraft in the scenario. Twenty pilots (64.5%) stated that heading was the first thing considered, while 11 pilots (35.5%) chose range. Pilot comments included:

"Heading aspect is most important between these two options. Also, altitude and trend are important considerations. Ultimately, if our flight paths will cross, then I am more likely to be concerned with a conflict, regardless of range."

"Based on range, you could rule many out quickly, regardless of climb/descent rate or speed. From there, the heading of the relevant aircraft ultimately highlighted the aircraft that posed true threats."

"Range was most important to those aircraft within several thousand feet. However the factors of descending or climbing in regards to heading are very important as well, making both very equal considerations in regards to converging traffic."

The pilots were also asked to rank order the importance of the secondary flight information given in the inset box for each aircraft (choices: aircraft's identification (ID), aircraft type, altitude, speed, course, and vertical trend; ranking of 1 being most important; 6 being least important). The results are shown in the box plot in Figure 2 showing the central tendency (mean rating (circle) and median), 25th and 75th percentile by the box height, 1.5 times the interquartile range by the whiskers, and asterisks denote outliers. Altitude was ranked of highest importance by 22 pilots (71%). The rankings of course and vertical trend were not statistically significant (T-Value = -1.19, P-Value = 0.245) enough to distinguish between a ranking of two or three. However, pilots generally agreed that speed, aircraft type, and aircraft ID were of lesser importance with rankings of four, five and six, respectively. Eleven pilots were in agreement as to a ranking order of: altitude, vertical trend, course, speed, aircraft type, and aircraft ID. This left 20 pilots choosing another nine separate ranking orders.

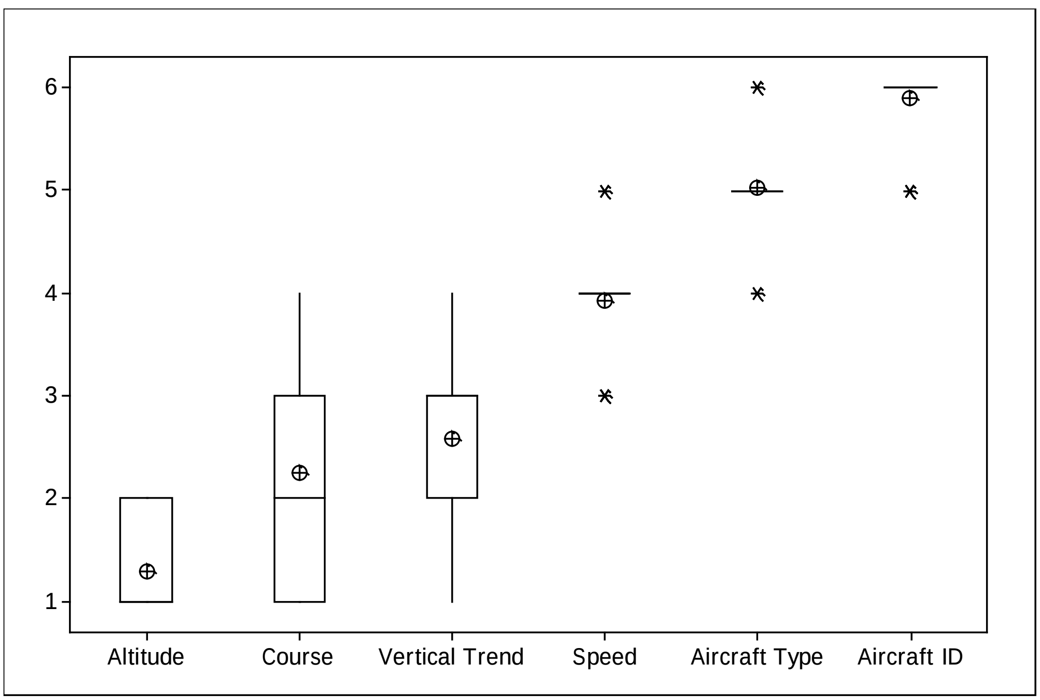

Figure 2. Rank Ordering of Secondary Aircraft Information (1: Most Important)

The pilots were asked to fill out a System Usability Scale (SUS) and ranked D.A.T.A.'s usability at 85.75 (Best). The SUS uses a five-point Likert scale (from Strongly Agree to Strongly Disagree) for a 10-item questionnaire. The scoring ranges from 0-100 and is seen as a reliable measure of usability. (Brooke, 1996) Pilots were also encouraged to give additional feedback of the system, if they wanted. Several pilots suggested the addition of ownship vertical speed to the data collection application, as they find it an important component when deciding relevancy. Others commented that knowing the selected aircraft's final climb or descent altitude would be helpful in determining the difference between "Maybe Relevant" and "Relevant/Not Relevant" These suggestions will be addressed in future work.

Discussion

We have now collected over 22,000 data points, from 31 ATPs, of traffic state data relative to ownship (i.e., altitude, course, range, bearing, speed) with pilot-derived relevancy labels (Relevant, Maybe relevant, and Not Relevant) to ownship. In addition, we have pilot-reported data on whether heading or range was the first consideration when choosing an aircraft to select. Pilots also ranked the order to which they used the aircraft information given to them when considering relevancy to ownship. After data collection, any discussion and additional pilot comments were also captured. These data were used to train machine learning algorithms that are designed to mimic human experts and their expertise. The detailed results of this work are reported elsewhere (Houston, Le Vie, in press) but overall, the algorithms are showing between 70% and 80% classification accuracy to the training. The results look promising, but not without challenges. One challenge experienced was making sure the machine learning algorithm was given not only enough data to train, test and learn from, but also enough of the "right" data. Through talking with pilots, a better understanding was gained in how they used the aircraft state data shown to make assumptions and predictions, which added extra information to their decision making. This was critical information that had not been considered and was not being provided to the machine learning algorithm. Having an expert walk through their decision-making process and selection method was fundamental in making sure that all of the processes that go into making a decision on whether an aircraft was relevant or not was captured and included. This effort continues to be a work-in-progress and is being used as a learning platform for the researchers to better gather this expertise in the future.

Future Work

This paper describes a data collection effort for training machine learning algorithms that will determine traffic relevancy, as the first-step in developing an intelligent party-line application and as a testbed for IAS development and evaluation. The TDM application is now running, using these training data, and shows promise in creating an intelligent decluttering and parsing agent.

The next immediate step is to assess the accuracy of the training data to the "expert" pilot population in general. An algorithm may never perfectly match the relevancy rating of every user. In fact, there is frequent disagreement among expert users about the "threat" of any individual aircraft. (St. John, Smallman, Manes, Feher & Morrison, 2005) In a study of six teams evaluating a threat management display, the interest level of an aircraft was agreed upon for only 41% of the aircraft. (Marshall, Christensen, & McAllister, 1996); Additional data collection up-coming will assess this training data against a new pilot population and assess the robustness of capturing expert pilot data for IAS development.

Future efforts will include real-time evaluation of the TDM algorithm performance and its ability to accurately predict air traffic relevancy in reference to ownship, the latency of its predictions, and its integration with other technologies being developed. As a learning/adaptive system, the stability of the algorithm as it adapts to the environment and changes will be assessed as this behavior may be a critical element in trust and human-autonomy teaming. Further, the system will be used for metrics development and evaluation as a Data Comm environment tool, assessing if relevancy changes as the operational context changes. The labels or markers for relevancy will also be expanded to include contextual or communication markers such as airport or runway identifiers, company names, routes of flight, etc.

Acknowledgments

This research was funded by NASA's Aeronautics Airspace Operations and Safety Program's Safe Autonomous Systems Operations (SASO) project. The contributions of Jason Neuhaus and Randall Bailey (NASA LaRC) were especially critical to the success of this work.

References

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability Evaluation in Industry.* London: Taylor and Francis.

DeCoursey, W. J. (2003) *Statistics and probability for engineering applications: With Microsoft Excel.* Boston, MA: Newnes.

Hastie, T., Tibshirani, R., Friedman, J. (2009) *The elements of statistical learning: Data mining, inference, and prediction, second edition.* New York: Springer-Verlag. doi: 10.1007/978-0-387-84858-7.

Houston, V. E. & Le Vie, L. R. (in press). *Autonomous System Technologies for Resilient Airspace Operations*

Marshall, S.P., Christensen, S.E., & McAllister, J.A. (1996). Cognitive differences in tactical decision making. In *Proceedings of the 1996 Command and Control Research and Technology Symposium* (pp. 122-132). Monterey, CA: Naval Postgraduate School.

National Research Council (2014). *Autonomy Research for Civil Aviation: Toward a New Era of Flight.* Washington D.C.: National Academies Press. doi: https://doi.org/10.17226/18815.

St. John, M., Smallman, H. S., Manes, D. I., Feher, B. A., & Morrison, J. G. (2005). Heuristic Automation for Decluttering Tactical Displays. *Human Factors*, *47,* 509-525. doi: https://doi.org/10.1518/001872005774860014

United States Air Force. (2005). Heuristic Automation for Decluttering Tactical Displays. *Human Factors*, *47,* 509-525. doi: https://doi.org/10.1518/001872005774860014

United States Air Force, "Technology Horizons - A Vision for Air Force Science & Technology During 2010-2030" AF/ST-TR-10-01, 15 May 2010.