Balancing transparency and conformance of an AI conflict detection and resolution support system

Carl Westin
Linköping University
Linköping, Sweden

Tiago Monteiro Nunes
Delft University of Technology
Delft, the Netherlands

Clark Borst
Delft University of Technology
Delft, the Netherlands

Stefano Bonelli
Deep Blue, Italy

Despite the recent turmoil in aviation, the trend is towards an increase in the amount of aircraft and miles flown. This growth adds additional pressure to an already strained Air Traffic Management (ATM) system. Part of the solution to this problem comes from the implementation of advanced automation support assisting the air traffic controller. One of the main challenges is how to best design and introduce new forms of automation incorporating Artificial Intelligence (AI). Machine Learning (ML) offers the possibility for introducing automation support systems able to sense, learn, and act autonomously.

Challenges stem in part from the fact that ATC problems, such as conflict detection and resolution (CD&R), are often complex and ill-defined. ATC represents a large "solution space" in that it accommodates many successful strategies within the basic system constraints. As long as aircraft are kept separated, controllers can differ in their overall strategies and individual solution choices. Increasingly, however, as the capabilities of automation approach those of the human, and machines can more often take over many of the 'thinking' parts of jobs like ATC, we are forced to consider issues of trust, acceptance, and reliance that will, invariably, be brought up.

Advanced automation and AI systems tend to be 'opaque,' unintuitive, and difficult to understand. Research has approached this problem in two ways: by creating conformal systems that strive to solve problems in line with the individuals' preferences (i.e. personalized systems), or by providing explanations for the system's behavior or reasoning underlying a recommended action (i.e. automation transparency, explainable AI, and ML interpretability). In the emerging age of AI and ML, the MAHALO project asks two simple but profound questions:

1. to what extent should we design automation to match individual human behavior (i.e., strategic conformal)? and
2. to what extent must it be transparent to the human (i.e., explainable)?

Answering these questions is important for the development of a human factors framework to guide design of future machine learning systems in safety critical domains. MAHALO focuses on conflict advisory (or decision support) automation, capable of providing the controller real-time assistance with CD&R. Thereto, MAHALO will develop an AI agent capable of detecting and resolving conflicts. The future reference environment is in line with the digital European sky vision 2040 (i.e. phase D in the digital transformation) in the European ATM Master plan. It is assumed that controller and automation collaborate in CD&R, with the automation conducting conflict detection and either executing solutions or proposing solutions, to the controller, to be accepted.

To balance the degree of conformance and transparency, we propose a hybrid ML approach that combines model-driven (e.g., rule-based expert systems) and data-driven (e.g. neural nets and reinforcement learning) applications. Such system could strive for optimal performance while accommodating individual differences. By knowing the individual's preferences, such system could provide transparency by explaining both why it suggests another solution (that deviates from the individual's), and why this solution is considered to be better.

The first MAHALO activity consisted of conducting a literature review to determine the state of the art in AI/ML approaches for CD&R. We classified the ML methods for CD&R according to an eleven-category framework. The categories were: problem addressed; avoidance planning horizon; target domain; airspace dimensions considered; main AI approach; multi-agent systems; feature engineering; resolution maneuvers; resolution objectives; conformance; and transparency. The framework is used as a guide to determine the most suitable method(s) for developing a ML CD&R system.

Four research clusters where found to be responsible for a majority of the research reviewed (11 out of 18 articles). The majority of research has focused on the conflict resolution problem. ML methods specifically for conflict detection have only been considered in three studies. The majority of previous approaches have restricted CD&R to a 2-dimensional representation of the environment (aircraft fixed to one altitude) and limited resolutions maneuvers to heading changes. Supervised learning approaches are most commonly used for detecting conflicts. For MAHALO, the most desirable ML method for the conflict resolution is likely to be a combination of a supervised learning algorithm together with a reinforcement learning algorithm. Supervised learning algorithms are particularly suitable for creating conformal systems that learning from expert knowledge, whereas reinforcement learning is suitable for finding more optimal solutions based on conflicting objectives.