Game Theoretic Modeling of Human Behavior in Mid-Air Encounters

Ritchie Lee
Carnegie Mellon University (SV)
NASA Ames Research Park
Mail Stop 23-11
Moffett Field, CA 94035
ritchie.lee@sv.cmu.edu

David H. Wolpert
Intelligent Systems Division
NASA Ames Research Center
Mail Stop 269-1
Moffett Field, CA 94035
david.h.wolpert@nasa.gov

Abstract

This paper introduces a novel modeling framework for hybrid systems involving automation and multiple human decision-makers. This framework, called "network-form games", combines Bayes nets with game theoretic concepts for modeling bounded rational humans. It allows the prediction, control and design of such hybrid systems. We illustrate the framework for automated recommender systems that help human pilots avoid mid-air collisions, presenting some parameter trade analyses of such systems.

1 Introduction

Traffic Alert and Collision Avoidance System (TCAS) is an aircraft collision avoidance system currently mandated on all US domestic aircraft with maximum take-off mass exceeding 5700 kg. It is an onboard system designed to operate independent of ground-based air traffic management systems to serve as the last layer of safety in the prevention of mid-air collisions (MACs). TCAS continuously monitors the airspace around an aircraft and warns pilots of potential threats by issuing a recommended action, or Resolution Advisory (RA), to the pilot.

While TCAS has performed satisfactorily in the past, one key weakness is that it assumes a deterministic pilot model. Specifically, TCAS assumes that a pilot receiving a TCAS RA will delay for 5 seconds, and then accelerate at 1/4 g to perform the RA maneuver. Although pilots are trained to obey TCAS RAs in this manner, a recent study of the Boston Area [1] has found that only 13% of RAs are fully obeyed -- the aircraft response maneuver met the TCAS design assumptions in vertical speed and promptness. In 64% of the cases, pilots were in partial compliance -- the aircraft moved in the correct direction, but did not move as promptly or as aggressively as instructed. Shockingly, the study also found that in 23% of the RAs, the pilots actually responded by maneuvering the aircraft in the opposite direction of that recommended by TCAS (although a number of these cases of non-compliance may be due to visual flight rules). The assumption is clearly not valid in the real world, and the associated risks of using the system will increase significantly in NextGen operation as air traffic density experiences exponential growth [2].

Pilot interviews have offered many insights toward understanding these statistics. During a mid-air encounter, the pilot does not blindly execute the RA maneuver. Instead, he combines the RA with other sources of information to predict his best course of action. In doing this, he quantifies the quality of a course of action in terms of a utility function defined over possible results of that course of action. That utility function does not only involve proximity to the other aircraft in the encounter, but also involves how drastic a maneuver the pilot makes. For example, if the pilot believes that a collision is unlikely based on his observations, he may opt to ignore the alarm and continue on his current course, thereby avoiding any loss of utility incurred by maneuvering. This is why a pilot will rationally decide to ignore alarms with a high probability of being false.

When designing TCAS, a high false alarm rate need not be bad in and of itself. Rather what is bad is a high false alarm rate combined with a pilot's utility function to result in pilot behavior which does not maximize expected social welfare. This more nuanced perspective allows far more powerful and flexible design of alarm systems than simply worrying about receiver operating characteristic (ROC) curves.

Here we elaborate this perspective. We introduce a framework for predicting the behavior of a hybrid system comprising automation and humans who are motivated by utility functions and anticipation of one another's behavior. We illustrate this framework by modeling the TCAS hybrid system. A fully formal elaboration of this framework, called "network-form games," can be found in [3].

2 Network-form game

A Bayes net can be used to model the behavior of a hybrid system involving only a single human by considering the human as a node in the network. (For example, this is the basis of influence diagrams.) Such a model reduces the human to a fixed conditional probability distribution. However, when there are multiple people in the system, more complex behavior arises because the humans determine their conditional distributions by anticipating the distributions of the other humans in the system. Such determination of conditional distributions by anticipating other conditional distributions is not considered in the conventional Bayes net literature [4]-[6]. However, it forms the core of game theory [7]-[9]. In particular, recent behavioral game theory models that replace the Nash equilibrium concept, such as the Quantal Response Equilibrium and Level-k Thinking, have been found to accurately predict results of experiments involving multiple interacting human behaviors [9]-[10].

Building on earlier approaches [5], in this paper we introduce a novel framework, "network form games," that combines Bayes nets and Game Theory to model hybrid systems that involve both automation and multiple interacting humans. In a net-form game, the Bayes net serves as the underlying probabilistic framework to model the system. Humans are quantified as special "player" nodes in the Bayes net, all other nodes being "nature nodes."

Formally, the distinction between nature nodes and player nodes is that nature nodes come with conditional probability distributions, while player nodes do not. Instead each player node is associated with a utility function, which maps the joint value of all the variables in the net to a real number. To fully specify the Bayes net it is necessary to determine the conditional distributions at the player nodes to go with the distributions at the nature nodes. The distribution at each player node is determined by using the set of *all* utility functions, via game theoretic equilibrium concepts. Here we do not formalize this in its full generality. Instead, in the next section we illustrate it for a particular equilibrium concept, on the aircraft collision avoidance problem.

3 Using a network-form game to model mid-air encounters

The complete net-form game representation for a three aircraft encounter is shown in Figure 1. At any time t, the true system state of the mid-air encounter is represented by the world state, S, which includes the dynamic states of all aircraft. Since the pilots and TCAS hardware are not able to observe the world state perfectly, a layer of nodes is introduced to model observational noise and incomplete information. W represents the pilot's observation of the world state, while W_{TCAS} represents what the TCAS sensors perceive. This model assumes that all aircraft are TCAS equipped, and W and W_{TCAS} are partial observations of S corrupted with Gaussian noise. A simplified model of the current TCAS logic is applied to W_{TCAS} to produce a TCAS RA, T. The player gets to observe W and T and chooses an aircraft command, π. Aircraft dynamics are simulated forward in time to the next time step. RAs of each aircraft are broadcasted due to the coordination mechanism of TCAS. For this reason, the RAs are also propagated to the next time step.

For computational simplicity, we assume that pilots only get to choose a single move, and they do so when they receive their initial TCAS RA. Furthermore, they do not change their move for the remainder of the encounter. When pilots are deciding their move, they assume that they are playing a simultaneous move game with other pilots. (Note that the timing of decisions is in reality stochastic as well as asynchronous. Adapting the framework to allow for such stochastic time is subject of ongoing research.)

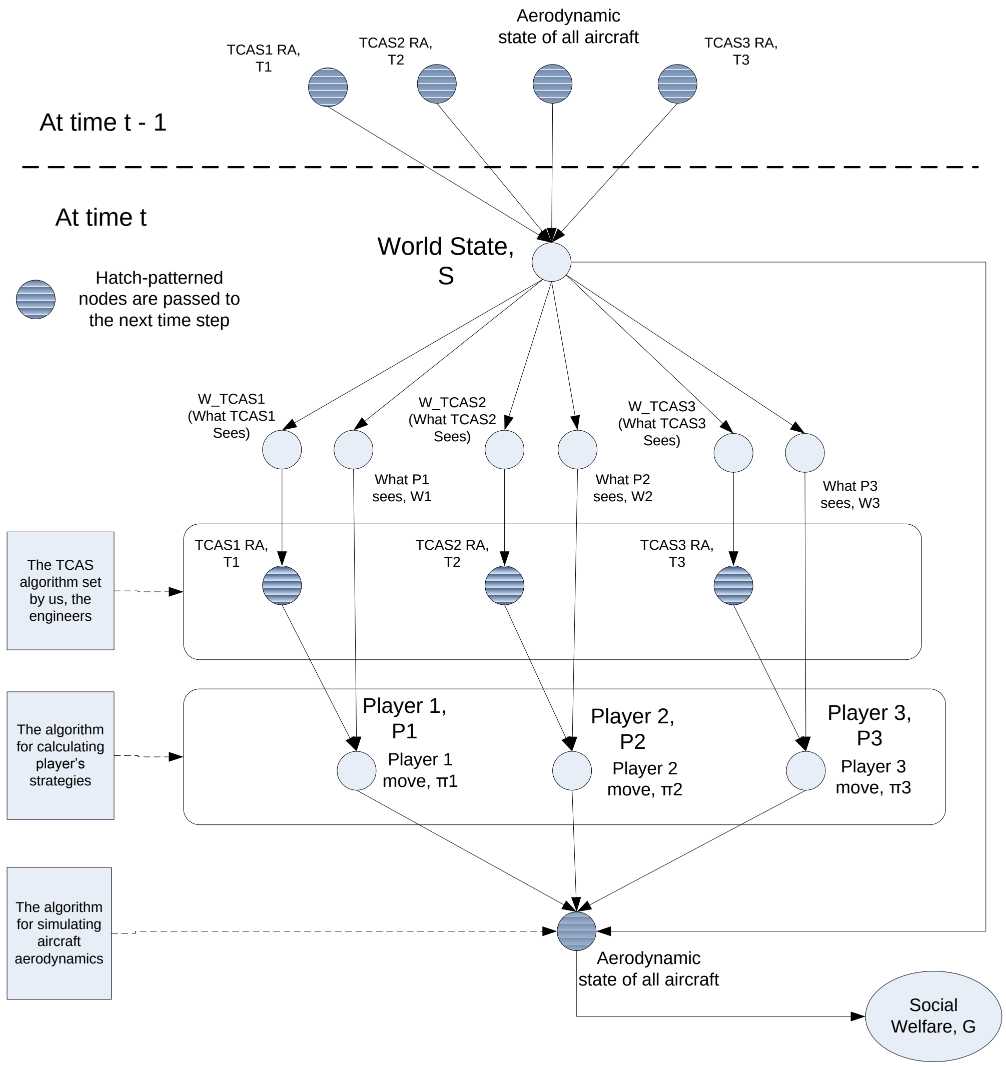

Figure 1: Network-form game representation of a mid-air encounter

We crafted player utility functions through interviews with pilots. (Refining these utility functions using experimental data is the subject of ongoing future work.) The player's utility function is modeled to be of the following form:

$$u_i = a_1 \log(\delta + d_{min}) - a_2|\pi_0 - \pi| - a_3\,|T - \pi|$$

where a_1, a_2, and a_3 are constant weights, δ is a small positive constant, d_{min} is the forecasted distance of closest approach between the aircraft, and π_0 is the pilot's current command. The utility function is a linear combination of three terms. The first term, which has the heaviest weight, reflects the pilot's fear of a collision by decreasing utility as d_{min} decreases. The choice of a logarithm function captures the fact that utility is much more sensitive when aircraft are close together than when they are far apart. The second term represents the pilot's desire to continue on his current trajectory, and thus penalizes moves that differ significantly from the current path. All other things being equal, the third term reflects an inclination to follow protocol, and thus to penalize moves that differ significantly from the issued RA.

Since d_{min} depends on the trajectories of all aircraft, which depend on the joint moves of all the players, each pilot must anticipate the other pilots' moves to decide how to act. However in doing this, the pilot realizes that the other pilot will be anticipating *their* move, etc., etc. To reduce this cycling to a prediction of pilot behavior we combine two experimentally validated game theoretic models, Level-k Thinking and Satisficing. We call this combination "Level-k Best-of-M."

A number of studies [7]-[10] have shown that Level-k Thinking performs well at predicting experimental data in games. The concept of Level-k is defined recursively as follows: A level k player plays as though all other players are playing at level k-1, who, in turn, plays as though all other players are playing at level k-2, etc. The process continues until level 0 is reached, where the player plays according to a prior distribution. Note that this may cause ricocheting to occur. For example, if player A is a level 2 player, he plays as though player B is a level 1 player, who in turn plays as though player A is a level 0 player. In most games, k is a fairly low number for humans; experimental studies have found k to be somewhere between 1 and 2 [7]. For simplicity, here we model all players as being level 2.

The concept of Satisficing [11]-[13] states that humans are unable to evaluate the probability of all outcomes with sufficient precision, and thus often make decisions based on adequacy rather than true optimum. The Best-of-M algorithm models this notion in the decision-making process as follows: The player samples M own candidate moves, and evaluates the expected value of each against M' sampled possible environment scenarios (i.e. the player considers M' samples of the joint probability of all nodes in the net minus those observed and those controlled by the player), and chooses to execute the move that results in the highest expected value of utility. M and M' are parameters that model player rationality, and are chosen to be 5 and 10, respectively in our studies except where indicated otherwise.

Note that to apply any game theoretic concept (not just level k) to set the conditional distribution at a player node, any unknown quantities in the utility function of that node must be inferred from values of the inputs to that node. (This reflects that humans in the hybrid system must estimate those quantities that they do not know but are concerned about.) In the conflict resolution problem, this means that players must estimate S, the other players' moves, and other unknown quantities given W and T only. The expected value of the player's utility function for a candidate move is given by:

$$
\begin{aligned}
&E(u_i \mid \pi_i, W_i, T_i) \\
&= \int dS' dW'_{TCAS_i} dWT'_{-i} d\pi'_{-i} P(S', W'_{TCAS_i}, WT'_{-i}, \pi'_{-i} \mid W_i, T_i) u_i(S', \pi_i, \pi'_{-i}, W_i, T_i) \\
&= \int dS' dW'_{TCAS_i} dWT'_{-i} d\pi'_{-i} Q(S', W'_{TCAS_i}, WT'_{-i}, \pi'_{-i}) u_i(S', \pi_i, \pi'_{-i}, W_i, T_i) \\
&\qquad\qquad\qquad \times \frac{P(S', W'_{TCAS_i}, WT'_{-i}, \pi'_{-i} \mid W_i, T_i)}{Q(S', W'_{TCAS_i},\ WT'_{-i}, \pi'_{-i})}
\end{aligned}
$$

where $Q(S, W'_{TCAS_i}, WT'_{-i}, \pi'_{-i}) \propto Q(W'_{TCAS_i} \mid W_{TCAS_i}) Q(S' \mid S) P(WT'_{-i} \mid S')\, P(\pi'_{-i} \mid WT'_{-i})$

is the importance sampling proposal distribution for variance reduction, and u_i is the utility function. Index i indicates self, whereas index –i signifies all players other than self. Substituting and expanding via Bayesian inversion, we have:

$$\propto \int dS' dW'_{TCAS_i} dWT'_{-i} d\pi'_{-i} Q(S', W'_{TCAS_i}, WT'_{-i}, \pi'_{-i}) u_i(S', \pi_i, \pi'_{-i}, W_i, T_i)$$
$$\times \frac{P(W_i, T_i \mid S')\, P(S') P(\pi'_{-i} \mid WT'_{-i}) P(WT'_{-i} \mid S')}{Q(W'_{TCAS_i} \mid W_{TCAS_i}) Q(S' \vert S) P(WT'_{-i} \mid S') P(\pi'_{-i} \mid WT'_{-i})}$$
$$= \int dS' dW'_{TCAS_i} dWT'_{-i} d\pi'_{-i} Q(S', W'_{TCAS_i}, WT'_{-i}, \pi'_{-i}) u_i(S', \pi_i, \pi'_{-i}, W_i, T_i)$$
$$\times \frac{P(W_i \mid S'_A) P(T_i \mid W'_{TCAS_i}) P(W'_{TCAS_i} \mid S') P(S')}{Q(W'_{TCAS_i} \mid W_{TCAS_i}) Q(S' \mid S)}$$

We approximate the integral above by sampling the distribution Q and calculating:

$$\approx \sum_{j=1}^{M'} \frac{P(W_i \mid S_A^j) P(T_i \mid W_{TCAS_i}^j) P(W_{TCAS_i}^j \mid S^j) P(S^j) u_i(S^j, \pi_i, \pi_{-i}^j, W_i, T_i)}{Q(W_{TCAS_i}^j \mid W_{TCAS_i}) Q(S' \mid S)}$$

Note that the prior distribution $P(S^j)$ is dependent upon the encounter model used; the proposal distributions $Q(W^j_{TCAS} \mid W_{TCAS})$ and $Q(S^j \mid S)$ are tight Gaussian distributions about W_{TCAS} and S respectively, and π_{-i} is calculated for one k level below. The players do not have access to the true quantities W_{TCAS} and S, but rather the simulator does, and therefore can make use of it as a technique to reduce variance.

4 Trade analyses

Because of its sampling nature, Level-k Best-of-M is well-suited for use with Monte Carlo techniques. In particular, we can use such techniques to assess performance of the overall system, using a social welfare function G defined as the minimum distance between aircraft during the encounter.

To demonstrate this capability, parameter trade analyses were performed on the mid-air encounter model, and sample results are shown in Figure 2. In each case, social welfare is observed while selected parameters are varied. The quantification of predicted improvements in social welfare (or any other metric) would be especially relevant to, for example, a system designer designing an improvement to the system, or a funder who is allocating resources for development projects. In Figure 2a, M_w and M_{WTCAS}, which are multiples on the noise of W and W_{TCAS} respectively, are plotted versus social welfare G. It can be seen that as the pilot and TCAS system's observations get noisier (e.g. due to fog or faulty sensors), social welfare decreases. However, a noteworthy observation is that social welfare decreases faster with M_w (i.e. when the pilot has a poor visual) than with M_{WTCAS} (i.e. noisy TCAS sensors). In Figure 2b, the dependence of social welfare on selected TCAS internal logic parameters DMOD and ZTHR are shown. These parameters are primarily used to define safety buffers around the aircraft, and therefore it makes intuitive sense to observe that there is an observed increase in G as these parameters are increased. Figure 2c plots player utility weights vs. social welfare. In general, the results agree with intuition that higher a_1 (i.e. stronger desire to avoid collision) and lower a_2 (i.e. weaker desire to stay on course) lead to higher social welfare (i.e. safety). This offers quantification on the potential increase in safety that additional training, regulations, incentives, and other pilot behavior-shaping programs may promise. Figure 2d plots model parameters M and M' versus G. These parameters are not ones that can be controlled, but rather ones that should be set as closely as possible to reflect reality. Note that model parameters in this study are uncalibrated, and are set according to the best judgment of the modelers. The primary focus of this project is to demonstrate the modeling technology, and thus a follow-on study to refine the model using experimental data is recommended.

Acknowledgements

We would like to thank Mykel Kochenderfer, and the Stanford Distributed Systems Control group for valuable feedback. We also thank the NASA IRAC project for funding this work.

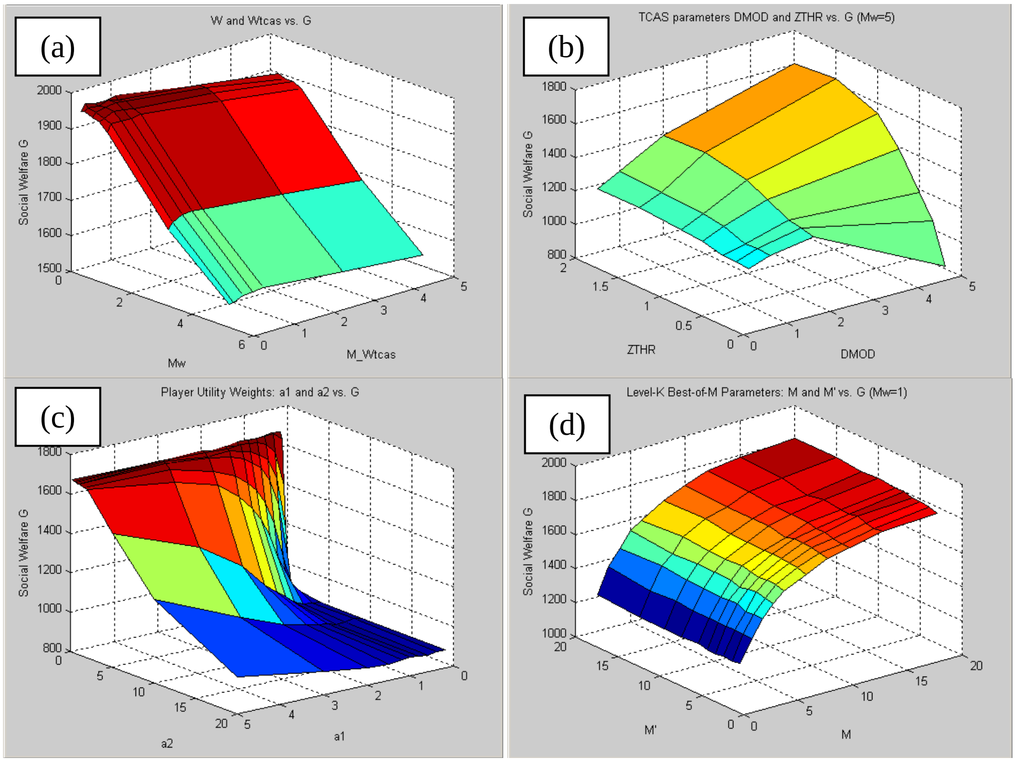

Figure 2: Trade analysis results of net-form game mid-air encounter model

References

[1] Kuchar, J.K. & Drumm, A.C. (2007) The Traffic Alert and Collision Avoidance System, *Lincoln Laboratory Journal*. (16)2.

[2] Federal Aviation Administration Press Release (2010) Forecast Links Aviation Activity and National Economic Growth. *http://www.faa.gov/news/press_releases/news_story.cfm?newsId=11232*

[3] Wolpert, D.H. & Lee, R. (2010) Network-Form Games. NASA ARC Working Paper.

[4] Koller, D. & Friedman, N. (2009) Probabilistic Graphical Models: Principles and Techniques. MIT Press.

[5] Koller, D. & Milch, B. (2003) Multi-Agent Influence Diagrams for Representing and Solving Games. *Games and Economic Behavior*. (45) 1, pp. 181-221

[6] Darwiche, A. (2009) Modeling and Reasoning with Bayesian Networks. Cambridge U. Press.

[7] Costa-Gomes, M.A., et al. (2009) Comparing Models of Strategic Thinking in Van Huyck, Battalio, and Beil's Coordination Games. *Journal of the European Economic Association*. 7.

[8] Camerer, C.F. (1989) An Experimental Test of Several Generalized Utility Theories. *Journal of Risk and Uncertainty*. 2(1), pp. 61-104.

[9] Camerer, C.F. (2003) *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

[10] Wright, J.R. & Leyton-Brown, K. (2010) Beyond Equilibrium: Predicting Human Behavior in Normal Form Games. *Twenty-Fourth Conference on Artificial Intelligence (AAAI-10)*.

[11] Caplin, A., Dean, M. & Martin, D. (2010) Search and Satisficing. NYU Working Paper.

[12] Simon, H.A. (1956) Rational Choice and The Structure Of The Environment. *Psychological Review*. 63(2), pp. 129-138.

[13] Simon, H.A. (1982) Models of Bounded Rationality. MIT Press.

[14] Kochenderfer, M.J. et al. (2010) Model-Based Optimization of Airborne Collision Avoidance Logic. *Massachusetts Institute of Technology, Lincoln Laboratory.* Project Report ATC-360.