Behaving Like Soldiers: A Multi-Agent System Approach to Course of Action Planning for Simulated Military Units

Chris Argenta, Rob Kewley, and Keith Brawner

Applied Research Associates, Inc., simlytics.cloud LLC, and US Army
cargenta@ara.com, rob@simlytics.cloud, and keith.w.brawner.civ@army.mil

Abstract

Training soldiers through virtual reality simulations has many benefits, two of which are reduced cost to conduct training operations and the ability to repeatedly perform activities that are not feasible in real life. The use of automated agents in place of human players or instructors has reduced cost. The sophistication of such agents can range from highly scripted scenarios designed to teach a specific concept to advanced artificially intelligent behaviors that react realistically to a wider range of in-game activities. Repeatability with realistic variation in the automated courses of action (COAs) is a primary objective, but often this comes at the cost of consistency to the doctrine (e.g., “what might I expect that support squad do in this case?”) and explainability of behaviors during after action review (e.g., “why did the enemy do that in this situation?”).

This paper describes early progress to develop advanced automated forces that are informed by a given corpus of doctrine. They anticipate future states with uncertainty to generate diverse COAs that guide higher level behaviors of nonplayer characters in a simulation, and they may be queried to explain behaviors at various points of the simulation. We are working with the US Army Combat Capabilities Command Soldier Center (CCDC-SC) to develop this technology for future integration into advanced virtual training systems.

Introduction: On the growing of the use of artificial intelligence in training simulations

The United States Army has invested in development of virtual training environments to support soldiers developing military skills. Simulations offer a method to train at a significantly lower cost per experience than live training as well as the ability to more easily tailor and repeat the training. However, effective training requires coordination within and between units, and interaction with friendly, neutral, and hostile entities. Experiential training in a multi-agent simulation environment can facilitate learning of military skills. However, managing large groups of human operators to train together (as in a massive multiplayer online games) is expensive, and challenging to control.

On the other end of the spectrum, entirely scripted non-player characters perform the roles of other units. With scripts, the lesson can be very well controlled but at the cost of diversity and variation of experiences. There are many approaches that fall within this range. For example, a trainer may manually but abstractly control a 9-person squad that moves and shoots as a unit. These strategies may use human players or scripts for higher level unit planning and artificial intelligence (AI) at the individual level to control the detailed and nuanced behaviors of automated soldiers.

In the research introduced with this paper, we seek even more affordable and flexible training by automating AI-driven agent behaviors for multiple higher-level units that coordinate or compete with live trainees. These agents are informed by a corpus of doctorial knowledge, so friendly, neutral, and adversary units plan and act consistent with training expectations but with significantly reduced scripting of scenarios. As a result, training experiences can be both wholistically educational and flexible enough to support variation and replay.

Additionally, we understand that significant conveyance of knowledge in training scenarios comes during after action review of the event by instructors. Intelligent automation can reduce the ability to explain the thinking underlying the behavior of automated forces. To address this challenge, we have prioritized unit-level automated reasoning that can be explained by way of linking it back to doctrinal decisions, a representation of the agent’s partial knowledge, and intentionally planned courses of action (COAs).

Mission Command Agents

Fundamentally, the US Army models soldier activities in distinct categories of move, shoot, communicate. Each trainee or non-player character conducts some combination of these actions, but unconstrained individual arena combat is better suited to a game than military training. Consider coordinating movement of a dismounted infantry squad by "flocking," a common game technique where agents adjust direction and speed based on simultaneous forces to maintain separation, stay together as a group, and move in generally the same direction (Fathy et. al. 2014). Flocking results in an emergent behavior that is not consistent how an infantry soldier in training should anticipate other squads to behave. US Army squads, for example, move 9 soldiers in formation with two identical fire teams of four and a squad leader moving with either element (HQDA 2016). On the other hand, Russian squads of seven personnel break down into a fire group and maneuver group of four and three soldiers respectively (Grau and Bartles 2016).

Executing behaviors and complex multi-agent maneuvers that are consistent with appropriate doctrine is critical for a training audience. So, before doing route planning for individuals, the planning algorithm must apply a doctrinal structure to constrain formations, spacing and movement techniques. This constrained action, using doctrinal templates to maintain explainable behaviors, must be applied by agents performing tasks at all organizational levels.

Our overarching planning architecture is based on the Army's Mission Command doctrine for command and control of Army forces (HQDA 2019). A Mission Command Agent is an intelligent automated system to develop mission orders for subordinate units. Figure 1 shows the Infantry Platoon Mission Command Agent. It receives a mission order that provides objectives in the form of a desired end state, tasks, intelligence about enemy forces and intentions, available resources, and constraints on action.

The situational understanding module aggregates this information with training simulation data, weather data, and terrain reasoning to update a local operational picture of relevant information. The visualize module continuously anticipates opposing (threat) courses of action, generates its own (friendly) courses of action, and stores them in the describe module. Doctrinal models aid in COA planning by performing doctrinal reasoning such as position analysis for different tasks. The assess module simulates COAs and assesses the degree to which friendly courses of action attain the desired end state. The direct module decides which COA to execute and passes orders to subordinate agents.

The complete agent architecture has a combination of different agents for control of different types of forces and different types of missions. For the example, the Infantry Platoon Mission Command Agent issues mission orders to subordinate Infantry Squad Mission Command Agents. This decomposition intentionally mimics the military organizational structures of the simulated units to ensure communication and reasoning is directly explainable in context of military doctrine. Each of these squads' Mission Command agents will execute their own modules to issue mission orders to their subordinate elements.

Towards Automated COA Planning

In our system, we extend methods of state-space planning to produce COAs that behave consistent with realistically given partial information and risk tolerances and consistent with the appropriate doctrine. Trainees should be able to probe an agent's behavior and find an explanation that provides useful insights into an agent's decision-making.

Multi-Agent Hierarchical Planning

When we compare tactical planning to other domains such as robotic soccer, there are similarities and differences. In robotic soccer, players play a simulated soccer game in accordance with their pre-trained AI policies, where the highest score within the allotted time wins (Kitano 1997). In such an environment, Texas' team applied the novel-at-the-time strategy of learning the simple actions within the environment prior to learning the complex ones – being referred to as layered learning (Stone and Veloso 1998). From optimized neural network policy logic at the low level (walking, running, dribbling), high-level behaviors were learned (ball interception), and, in turn, passing behaviors and positioning behaviors. The soccer playing agents in such a simulation rely upon their low-level actions to scale to team behaviors.

Our approach is building from this structure in developing the lower-level doctrinally correct actions prior to developing the higher-level strategy which determines the tactics. Our researchers have applied this approach previously in the military domain (Kewley and Embrechts 2008), but without the doctrine and explainability requirements.

One significant difference in our problem domain is that the objective is not simply to win a tactical engagement. Instead, we want simulated agents to behave reasonably for the scenario, reliably with respect to situationally applying doctrine, and reactively to the actions taken by the trainee. COA planning for organized forces is a multi-agent planning problem where agents exist within a hierarchical structure. We focus on centralized planning for agent teams, as opposed to distributed planning at each agent and communicating their plans, because training scenarios with automated friendly force partners usually specify higher echelon orders that are decomposed to subordinate units, made up of either automated agents or human trainees.

Multi-Objective / Multi-Solution Planning

Most AI planning systems find an optimal plan, but this is complicated when the utility function balances between multiple objectives. In our case, an example planning task determines a movement route to a goal position that must concurrently be feasible and appropriate for the mission; minimizes distance, time, and metabolic cost; reduces risk based on an uncertain understanding of enemy positions; and is consistent with the doctrine and the formations required. In a specific scenario there may not be a solution that maximizes all these objectives, so trade-offs must occur.

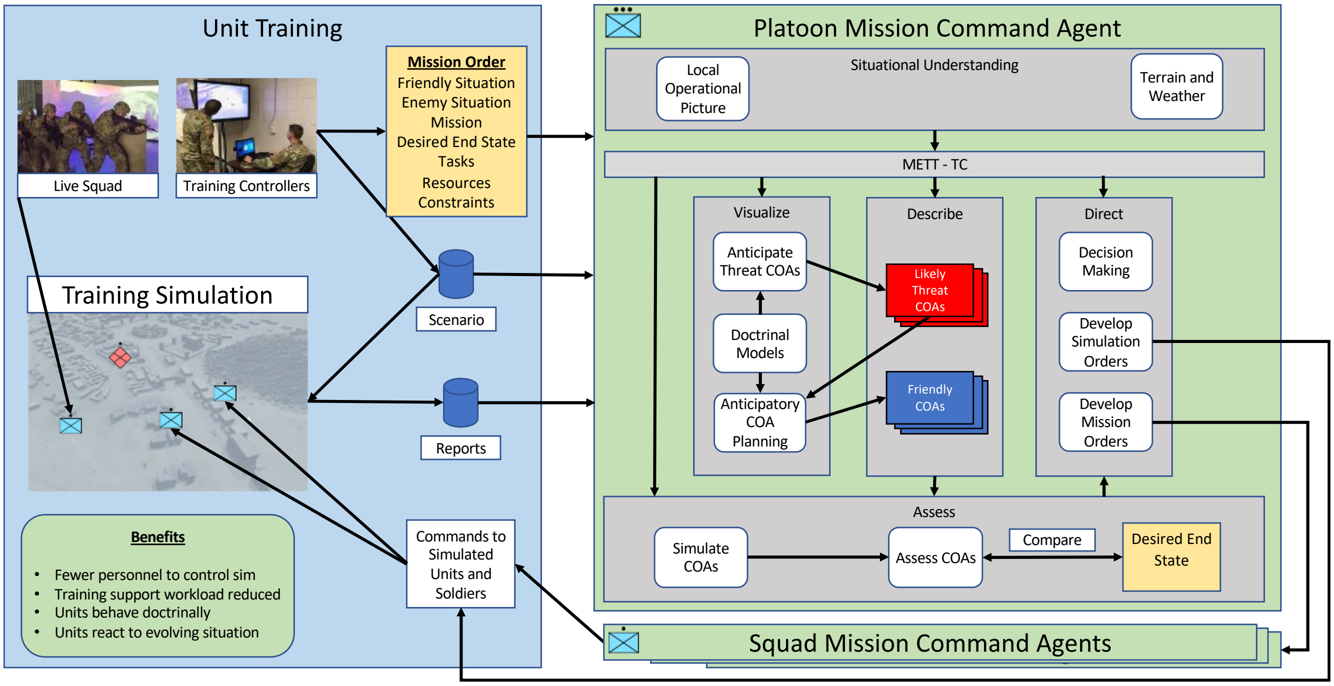

Figure 1. A Mission Command Agent for an infantry platoon playing the role of friendly forces.

While a common solution to multiple objectives is to give each objective a relative weight and aggregate them into a single utility function, our planner generates a set of plans along a Pareto frontier that corresponds to different trade-offs in objective weights and yields a set of plans that accomplish the goal in diverse ways. By evaluating the variation in the set of plans, we can determine which trade-offs have the most impact (Argenta, Hale, Doyle 2015).

In our approach, objectives such as mobility and shortest path follow traditional utility functions, for which heuristics exist to inform state-space planning. Other objectives such as consistency to doctrine, integrity of formation, and risk taking are informed by force-specific doctrinal models implemented as critics that evaluate the emerging plans akin to constraint satisfaction problems. Forgoing the requirement of finding a single optimal solution allows us to combine these strategies and better explain solutions by identifying the trade-offs across both quantitative (e.g., distance) and qualitative (e.g., consistency to doctrine) considerations, resulting in more realistic COAs.

Terrain Reasoning and Abstraction

Terrain plays a critical role in the behaviors of ground forces. Features such as ground covering, ridge lines, military crest, draws, clearings, and roads refer to areas with common characteristics and semantic meaning within doctrine that allows us to aggregate areas with similar features and treat them as single locations. Like bounding volume hierarchies, clustering these locations into a hierarchy provides us three benefits. First, planning at a high level of abstraction before filling in the details can improve state-space search algorithms. Second, using higher level abstracts to assess variation, we enforce meaningfully diverse alternatives. Finally, more abstract notions of areas (e.g., around valley Alpha) is consistent with using less precision in representing locations when predicting the movement of others with limited information.

Anticipatory Planning

Traditional planning often assumes full knowledge of the world. However, full knowledge is rarely the case in military operations, and the behavior of other agents and teams can be both difficult to predict and reactive to those of others. The best we can do is forecast future states and anticipate potential agent behaviors. Anticipatory thinking is the cognitive practice of considering potential futures and analyzing them to avoid surprise (Argenta 2019). We extend this to Anticipatory Planning by generating a range of diverse but feasible future states and trajectories for other agents, weighted by likelihood. Our future states are then used as uncertain proxies of full knowledge of the state.

The entity performing COA planning may observe some of the enemy's activity, from which it may infer the agent's organizational structure and its current plan. Plan recognition is the process of abductively reasoning about plan that underlies observed actions, and this process may result in multiple feasible interpretations of the available observations (Argenta, Doyle 2017). We use plan recognition in anticipatory planning to more effectively leverage available knowledge to determine the utility during planning. Planning under this uncertainty takes the form of considering the likelihood of a threat position when assessing the risk of a specific segment of a route and in developing contingencies in potential scenarios.

Planning Example

In our evaluation scenario, friendly forces occupy a support by fire positions based on the anticipated enemy positions as shown in Figure 2. Doctrine calls for maximizing fields of fire into enemy positions while also maintaining a stand-off distance up to 800 meters. Terrain reasoning aggregates possible locations into eight areas that represent military crests along a ridge line (green rectangles in Figure 2). The doctrinal model produces 3 different Pareto-optimal solutions, based on different weightings of the objectives. The optimal observation objective places it in Area A, which, at 200 meters, is too close to enemy forces. Area B, at 400 meters, provides the best stand-off, but has poor visibility of the threat. Using combined weights, Area C has the most utility, as indicated by the darkest green.

As a basis of comparison, a subject matter expert manually planed SBF 3 position for this scenario, selecting Area D. This area is 12% worse for standoff and 33% worse according to observation and fields of fire. This planning example demonstrates effective selection of tactical locations while also following an explainable doctrine.

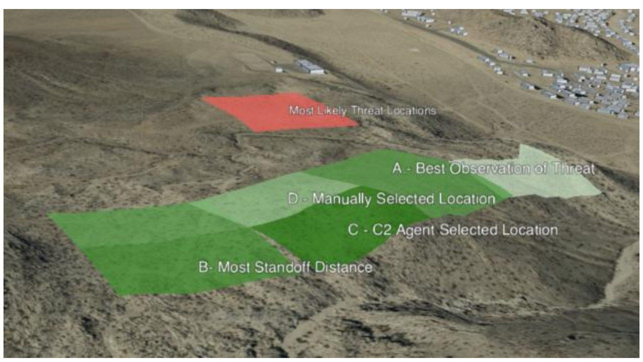

Figure 2. Friendly course of action for support by fire position to suppress anticipated threat positions (shaded red).

Conclusions and Future Work

While not complete at the time of this paper, we present an early example that demonstrates feasibility of several key capabilities in the context of a realistic training scenario. As our research progresses, we expect to be building on these early demonstrations to establish a more comprehensive corpus of doctrinal models. We will be creating additional mechanisms to automate reasoning over this corpus and applying this knowledge to anticipating what an enemy or other friendly force might do. Finally, our AI-enabled Mission Command Agent will reason over a larger set of feasible COAs using multiple objectives and trade-offs.

We are working with the US Army CCDC-SC to develop and integrate these capabilities into their advanced training and evaluation systems. During this project, we will develop simulation services for the evolving infrastructure, as well as additional interfaces for directing and explaining automated behaviors.

Acknowledgements

This material is based upon work supported by the US Army CCDC-SC, Soldier Effectiveness Directorate (SED), SFC Paul Ray Smith Simulation and Training Technology Center (STTC) under contract No. W912CG-21-C-0018. Any opinions, findings and conclusions, or recommendations expressed are those of the author(s) and do not necessarily reflect the views of the CCDC-SC-SED STTC.

References

Argenta, C., Browning, A., Amos-Binks, A., and Lyle, M., 2019. Avoiding Surprise: Augmenting Anticipatory Thinking with Scenario Explorer. *Cognitive Systems for Anticipatory Thinking, AAAI Fall Symposium,* Washington DC.

Argenta, C., Doyle, J., 2017. Probabilistic Multi-Agent Plan Recognition as Planning (P-MAPRAP): Recognizing Teams, Goals, and Plans from Action Sequences. *International Conference on Agents and Artificial Intelligence,* Porto Portugal.

Argenta, C., Hale, C., and Doyle, J., 2015. Assessing Variation of Adaptive Game-Based Training with Event Sequence Alignment and Clustering. *Advances in Cognitive Systems*, Atlanta GA.

Fathy, H., Raouf, O, and Abdelkader, H. 2014. Flocking Behavior of Group Movement in Real Strategy Games, *9th International Conference on INFOrmatics and Systems (INFOS2014).*

Grau, L. and Bartles, C, 2016. *The Russian Way of War: Force Structure, Tactics, and Modernization of the Russian Ground Forces.* Foreign Military Studies Office.

Headquarters, Department of the Army, 2016. *Infantry Platoon and Squad.* Washington D.C.: The Army Publishing Directorate.

Headquarters Department of the Army 2019, *Army Doctrine Publication 6-0 – Mission Command: Command and Control of Army Forces*. Washington, D.C.: The Army Publishing Directorate.

Kewley, R. and Embrechts, M 2008. A Multi-Agent System for Tactical Control of Automated Forces. *International Command and Control Journal* 2(2).

Kitano, H, et al. 1997. RoboCup: A challenge problem for AI. *AI magazine* 18(1): 73-73.

Stone, P. and Veloso, M. 1998. Layered approach to learning client behaviors in the robocup soccer server. *Applied Artificial Intelligence* 12(2-3): 165-188.