Explainable Artificial Intelligence for Training and Tutoring

H. Chad LANE, Mark G. CORE, Michael VAN LENT, Steve SOLOMON, Dave GOMBOC
University of Southern California / Institute for Creative Technologies
13274 Fiji Way, Marina del Rey, CA 90292 USA
{lane, core, vanlent, solomon, gomboc}@ict.usc.edu

Abstract This paper describes an Explainable Artificial Intelligence (XAI) tool that allows entities to answer questions about their activities within a tactical simulation. We show how XAI can be used to provide more meaningful after-action reviews and discuss ongoing work to integrate an intelligent tutor into the XAI framework.

Introduction

Military training aids typically provide an after-action review (AAR) tool to allow students to review their exercises and ideally learn from them. Common features of these tools include mission statistics, a list of accomplished and failed objectives, and sometimes a mission replay feature. Because of increasingly complex artificial intelligence (AI) in these training aids, it has been difficult for users of such AAR tools to understand how their orders translate into the activities of computer-controlled entities with such limited AAR tools. Student users have the additional disadvantage of possessing fragmented and possibly flawed domain knowledge: they are faced not only with learning new tactical knowledge (i.e., how units perform their tasks) and new skills (i.e., constructing plans and updating them on the fly), but also comprehending emergent behaviors and their triggers.

To provide a better AAR tool and to help users better understand entities' actions in military simulations, we have developed a portable Explainable AI (XAI) module that allows a user to question entities directly about their actions, status, and goals. For live training exercises, the U.S. Army defines AARs as "a review of training that allows soldiers, leaders, and units to discover for themselves what happened during the training and why. It is also used to solicit ideas on how the training could have been performed better" [2, App. G]. During such AARs, participating soldiers are usually available for questioning. The goal of our research is to use XAI to provide this feature for computer-controlled soldiers.

1. Explainable Artificial Intelligence for Training

Although our current XAI system was not designed exclusively for pedagogical purposes, it nonetheless addresses the first part of the AAR definition given above: it allows for a deeper investigation of what happened during an exercise. To interact with entities, the student may select from time points of focus, a particular entity to question, and finally a question to ask. Some questions are specific to a particular entity (e.g., what is your health?) while others concern a larger group (e.g., what is your unit's task?). Time points and available questions are dependent on the entity in focus and the discourse context and exchanges with the system are displayed as question and answer dialogues. More details about the interface and system implementation can be found in [3].

1. User: What is your health/damage status?
2. Gligoric [2:26]: **I am unhurt.**
3. User: Let's fast forward to simulation time 4:16.
4. Gligoric [4:16]: **I am currently withdrawing.**
5. User: What is your health/damage status?
6. Gligoric [4:16] **I have been killed in action.**
7. User: Let's talk to Private Evans.
8. Evans [4:16]: **I am the fire team leader of Blue Fireteam 2. My unit's task is to clear the W room of the post office.**
9. User: Let's rewind to simulation time 2:32
10. Evans [2:32]: **I am currently transiting to the right stack position outside of the NW door of the W room of the post office.**
11. User: What are you shooting at?
12. Evans [2:32]: **I am shooting at an entity located at 125 m NNE of the NW door door of the W room of the post office.**

Figure 1. Sample XAI dialogue

A sample interaction with our XAI system appears in Figure 1. The student begins by asking questions of Gligoric, an opposing force entity (OPFOR). Later, the student changes to Evans, a blue force entity (BLUFOR). This dialogue is intended to show how a student might investigate an event using the XAI tool. The student here probably believes Evans' unit was somehow involved in the death of Gligoric and is examining the events leading up to it. Although some of this information could potentially be derived from a graphical replay (we have provided this capability in previous work), there is a chance it will not be obvious to a student from the replay (e.g., line 9). Such questioning can also therefore be used by the student to confirm inferences made from visual evidence alone. In its current state, XAI is best viewed as a *discovery environment*. That is, the student is free to use the tool as desired to explore the simulation results and achieve the goals of the AAR. Having just completed the exercise for "real," interactions, such as the one in the figure, give the student a chance to learn more about the entities and what they experienced. It is up to the student, however, to ask the right questions of the right entities and understand the responses.

Focusing more specifically on our system's dialogue manager and natural language generator, we see that pedagogical support is built into these components. Currently we maintain a simple dialogue state consisting of all the entities and units that the user has talked with. In the dialogue in figure 1, Evans introduces himself as fire team leader and describes his unit's task because the student has not talked with either Evans or anyone else in that unit. This feature is a placeholder for more powerful reasoning about how to adapt the system's output to the student (e.g., it should not use undefined technical terms, it may need to explicitly state knowledge implied by its explanations). Although it is currently simulation-dependent, our system also maintains specific points of reference to refer to when responding to questions that require some location-oriented answer (e.g., line 12 in the Figure 1).

2. Related Work

The motivation for and technical challenges of explaining the internal processing of AI systems have been explored at length in the context of medical diagnosis systems. One prominent example, MYCIN, used a complex set of rules to diagnose illness and suggest treatments based on patient statistics and test results [6]. The developers of these systems were quick to realize that doctors were not going to accept the expert system's diagnoses on faith. Consequently, these systems were augmented with the ability to provide explanations to justify their diagnoses. Education becomes a natural extension as well since explanation is often an important component of remedial interventions with students. Three notable efforts falling into this category are the Program Enhancement Advisor (PEA) for teaching LISP programmers to improve their code [5], the family of successors to MYCIN [1], and another entity-driven explanation system, Debrief [4].

3. XAI for Tutoring

Evidence for learning in pure discovery environments is marginal [5], and so we are in the early stages of designing an intelligent tutoring module with the goal of providing a more guided discovery experience for students. We adopt the general goals of an AAR: review what happened, investigate how and why these events occurred, and discuss how to improve future performance. Answering *why* questions is a significant technological challenge, but also highly relevant to good tutoring. For example, discovering why a unit has paused in the middle of executing a task has the potential to help a student who gave the order to proceed. This may require reasoning about previous or concurrent events in the simulation. If a unit is currently under fire, for example, it is critical that the student understand what has caused the delay. It could very well involve an earlier mistake, such as failing to provide cover. The student could be asked to analyze the situation and suggest ways to allow the unit in question to proceed. One such question would be "Now that you have learned why this unit is delayed, what was missing from your plan?" If the student cannot generate any ideas, hints such as "Can you think of a way to conceal the unit for safe movement?" or "Do you see any other nearby units that could provide cover fire?" would be appropriate. We hypothesize that questions such as these, and more dynamic AARs, will improve students' self-evaluation skills and problem solving abilities within the simulation.

In addition to working with tactical behaviors, we are also in the early phases of targeting non-physical behaviors, such as emotional response, for explanation. This has advantages for systems that aim to teach subjects such as negotiation skills, cultural awareness or sensitivity. Explaining why an utterance (by a user) has offended an automated entity is, for example, similar to explaining emergent tactical behaviors. Tutoring in situations like this would, we believe, also be similar (e.g., "Could you have phrased that differently?").

Acknowledgements

The project or effort described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

[1] Clancey, W. J. (1986) From GUIDON to NEOMYCIN and HERACLES in twenty short lessons, *AI Magazine*, Volume 7, Number 3, pages 40-60.

[2] FM 25-101. (1990) Battle Focused Training. Headquarters, US Dept. of the Army. Washington D.C.

[3] Gomboc, D., Solomon, S., Core, M. G., Lane, H. C., van Lent, M. (2005) Design Recommendations to Support Automated Explanation and Tutoring. To appear in *Proceedings of the 2005 Conference on Behavior Representation in Modeling and Simulation (BRIMS)*, Universal City, CA. May 2005.

[4] Johnson, W. L. (1994) Agents that learn to explain themselves. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 1257-1263.

[5] Mayer, R. M. (2004) Should There Be a Three-Strikes Rule Against Pure Discovery Learning? *American Psychologist*, Volume 59, Number 1, pages 14-19.

[6] Shortliffe, E. H. (1976) Computer-based Medical Consultations: MYCIN. Elsevier, New York.

[7] Swartout, W. R., Paris, C. L., and Moore, J. D. (1994) Design For Explainable Expert Systems. *IEEE Expert*, Volume 6, Number 3, pages 58-64.