Using virtual tour behavior to build dialogue models for training review

Antonio Roque*, Dusan Jan, Mark Core, and David Traum

USC Institute for Creative Technologies
Playa Vista, CA, USA
aroque@ucla.edu,{jan,core,traum}@ict.usc.edu

Abstract. We develop an intelligent agent that builds a user model of a learner during a tour of a virtual world. The user model is based on the learner's answers to questions during the tour. A dialogue model for a simulated instructor is tailored to the individual learner based upon this user model. We describe an evaluation to track system accuracy and user perceptions.

Keywords: User models and adaptive agents, Dialogue models

1 Introduction

Researchers have investigated the use of intelligent agents in virtual worlds to act as virtual guides [5, 6] which lead human-controlled avatars through a virtual environment for entertainment or instructional purposes. Human-agent dialogue is an important part of such a **virtual tour**, not only to provide information and coordinate movement, but also to play a role in instruction. One instructional role is ascertaining how well the human learner is understanding the information being presented.

Our e ort is motivated by research in educational systems which use dialogue strategies tailored to individual learners. Researchers have tracked learners' behavior to build user models during experiences such as interacting with multimedia learning environments to learn shipboard emergency management [10] or writing essays to answer qualitative physics problems [3, 4]. These user models then guide subsequent dialogues tailored to the individual learner's needs.

In our research we explore applying such a user modeling technique during a virtual tour. We track a learner's dialogue behavior during an instructional virtual tour and build a user model that is combined with a model of the teaching goals to build a dialogue model for a post-exercise discussion. This dialogue model is used by a virtual guide to hold a multi-channel discussion in which private messages for individual learners can be customized based on their dialogue model.

* Now at the University of California Los Angeles, Computer Science Department

2 Virtual Tour Testbed

Educators have demonstrated the value of virtual worlds as learning environments using a variety of platforms including Active Worlds [2, 7]. The testbed we adopted was created in Active Worlds by US government contractors, and included a Powerpoint-style briefing in the virtual world followed by a virtual tour of a roadblock in Iraq. The existing virtual tour was designed to be directed by a human guide.

We developed an intelligent agent to conduct the virtual tour when a human tour guide was not available. This **virtual guide** led the learners through the virtual world, communicating through recorded audio narration mirrored by on-screen transcriptions, answering learner-initiated questions regarding the scenario, and asking the learners a number of questions to gauge their understanding of the material.

We also added an After Action Review (**AAR**) [1] following the virtual tour, during which the virtual guide used an interactive question and answer format to reinforce the lessons it had previously discussed. We personalized this interaction by tailoring the style of interaction to the learner's needs as described below. During the AAR, the virtual guide interacted with the group of learners through a variety of channels: using text chat to the entire group to manage the AAR, and using private in-world Instant Messages with each individual learner to ask questions and deliver didactic content. The in-world Instant Messages allow the pace of the AAR to be tailored to the level of knowledge shown by an individual learner. Figure 1 shows an AAR in progress.

Fig. 1. An After-Action Review (AAR) being conducted

3 Models of Knowledge Components, Users, and Dialogues

To implement an automated AAR, we first needed a model of the information to be taught. Following [8], we use the concept, **knowledge component**, "an acquired unit of cognitive function or structure that can be inferred from performance on a set of related tasks" (p. 9). We developed a set of knowledge components for the domain by analyzing the background reading material as well as studying a recording of a virtual tour conducted by a human subject matter expert. Of the thirteen knowledge components for the domain, nine are covered by guide explanations during the virtual tour while four are covered by questions asked.

After posing a question during the virtual tour, the virtual guide waits briefly for replies. Learner responses are not individually prompted, praised, or corrected at that time; after a pause, the agent only provides the answer and continues on the tour. However, if the learners make responses to the questions then a statistical natural language intepreter [9] automatically classifies the responses as either "Right" or "Wrong", and those classifications are saved for use during the AAR.

Each of these questions is linked with a relevant knowledge component, and the learner's user model consists of variables representing whether the learner has demonstrated competence in these knowledge components by answering the associated question correctly. For each knowledge component, either the learner has demonstrated evidence of understanding of the material (by answering the related virtual tour question correctly), or they have demonstrated a possibly incorrect understanding (by answering the virtual tour question incorrectly), or they have provided no indication of their understanding (by not answering the question), or there is no evidence that the learner was even present when the question was asked. This last condition could occur if, for example, a learner had lagged behind in the tour.

Each knowledge component is linked with dialogue strategies for use during the AAR. The default dialogue strategy contains a question to test the learner's knowledge of the knowledge component and a statement of the correct answer. The guide gives positive feedback after correct answers and provides the statement of the correct answer after both correct and incorrect answers. A vague or incomplete answer may still be classified as correct so we show the pre-authored correct answer to avoid potential misconceptions.

For some knowledge components, an alternate dialogue strategy of simply summarizing the knowledge component is available. If the user model indicates that there is evidence that the learner has mastered this knowledge component, the guide will select the strategy of summarization during the AAR. Otherwise, the question-answer strategy will be used to test the learner's understanding.

For example, one knowledge component concerns the high levels of stress felt by the soldiers at the checkpoint due to previous attacks. The virtual tour question, which follows a description of the recent violence, is: "How do you suppose that a ects the soldiers at the checkpoint?" If the learner answers that question correctly during the virtual tour, then the virtual guide selects a "review" strategy during the AAR, and summarizes the concept. If the learner answers the virtual tour question incorrectly, then the virtual guide selects a "remediation" strategy during the AAR, first asking a question, optionally providing positive feedback if they answer the AAR question correctly, and concluding with a statement summarizing the correct answer. Examples of both strategies are shown below.

Once the AAR begins, the virtual guide's dialogue manager iterates through a queue of knowledge components. For each knowledge component, the virtual guide queries the user model to pick a dialogue strategy to address the knowledge component.

A sample dialogue is shown in Table 1. In line 1, the virtual guide asks a question during the virtual tour that in line 2 the learner answers correctly. Line 3 occurs later in the virtual tour and corresponds to the virtual guide asking another question although in this case the learner's answer in line 4 is vague enough to be classified as incorrect.

#	Speaker	Text
1	Virtual Guide	Having said that, how do you suppose that a ects the soldiers at the checkpoint?
2	Learner	they must be scared to be in danger (Classified as: Right Answer)
...		
3	Virtual Guide	What inherent dangers of the checkpoint do you see?
4	Learner	it's open (Classified as: Wrong Answer)
...		
5	Virtual Guide	Given this situation [of frequent terrorist attacks], the soldiers at the checkpoint were likely tense especially since they had experience there and had some people die.
6	Virtual Guide	What were the specific dangers of the checkpoint due to the surrounding terrain?
7	Learner	the overpass
8	Virtual Guide	Right.
9	Virtual Guide	The soldiers at the checkpoint have no cover from attacks originating from the overpass or the nearby buildings...

Table 1. Sample Dialogue

Line 5 shows the discussion, in the AAR, of the same knowledge component addressed in lines 1 and 2. The virtual guide chooses a "review" strategy for the knowledge component given that the learner answered the question in line 1 correctly. Lines 6-9 address the same knowledge component addressed in lines 3 and 4. In contrast with the previous example, the virtual guide uses a "remediation" strategy given the the learner answered the question in line 3 incorrectly.

Because the learner answered correctly in line 7, they are given positive feedback as well as the pre-authored version of the correct answer.

4 Evaluation

Two major question about the e ectivness of the AAR are: does it result in learning gains beyond those provided by just the virtual tour, and does it make the experience better or worse from the learner's perspective? We did not have a test to evaluate learner knowledge of the domain before and after the experience so we focused solely on the performance of the classifier on learner answers, and learner evaluations of the experience.

The accuracy of the statistical natural language interpreter heavily influences the accuracy of the dialogue model and whether learners receive an appropriately customized AAR and appropriate positive feedback. Learners may also get frustrated if they perceive the system as not understanding them.

To evaluate the system, we ran several full sessions including the initial briefing, the virtual tour, and the AAR. Three of these sessions (10 data points) were used to evaluate the statistical natural language interpreter. Two human annotators labeled the correctness of each learner reply made during the virtual tour, and compared this to the automated classifications. On the 10 data points, the inter-annotator reliability as measured by Kappa was 0.79. The automated classifications agreed with human consensus classifications 80% of the time, which means the majority of time the AAR was correctly tailored to the individual learner.

In addition to recording the system behavior, we also collected qualitative data with post-session surveys in four of the sessions; responses on a seven-point Likert scale showed above-average scores in questions related to natural language understanding during the AAR, the AAR experience and the experience as a whole, as shown in Table 2.

Question	**Learner 1**	**Learner 2**	**Learner 3**	**Learner 4**
The After Action Review improved the experience. (1=Not at all, 4=Somewhat, 7=Very much)	4	7	6	6
How well do you think the AAR bot understood you? (1=Not at all, 4=Somewhat, 7=Perfectly)	4	5	6	5
How did you like the [General] experience? (1=I disliked it very much, 4=Neither liked nor disliked, 7=I liked it very much)	5	7	5	6

Table 2. Responses to Post-Session Questionnaire

5 Future Work

Future work will involve more extensive testing of this approach either in the domain discussed above or in other domains. Developing a pre- and post-test for domain knowledge will help better evaluate the system and judge the impact of potential new features such as a more detailed user model and more sophisticated AAR dialogue strategies.

Acknowledgments

This work 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. Army, U.S.: A leader's guide to after-action reviews (September 1993), TC 25-50
2. Dickey, M.D.: Three-dimensional virtual worlds and distance learning: two case studies of active worlds as a medium for distance education. British Journal of Educational Technology 36(3), 439–451 (May 2005)
3. Forbes-Riley, K., Litman, D.: Adapting to student uncertainty improves tutoring dialogues. In: Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED) (July 2009)
4. Forbes-Riley, K., Litman, D.: Using performance trajectories to analyze the immediate impact of user state misclassification in an adaptive spoken dialogue system. In: Proceedings 12th Annual SIGdial Meeting on Discourse and Dialogue (SIGdial) (June 2011)
5. Gayle, R., Manocha, D.: Navigating virtual agents in online virtual worlds. In: Proceedings of Web3D 2008, the thirteenth Symposium on 3D Web Technologies. pp. 53–56
6. Jan, D., Roque, A., Leuski, A., Morie, J., Traum, D.: A virtual tour guide for virtual worlds. In: Proceedings of Intelligent Virtual Agents 9th International Conference (2009)
7. Jarmon, L., Lim, K.Y., II, B.S.C. (eds.): Journal of Virtual Worlds Research, vol. 2 (April 2009), Special Issue on Pedagogy, Education and Innovation in Virtual Worlds
8. Koedinger, K.R., Corbett, A.T., Perfetti, C., the current, past members of the Pittsburgh Science of Learning Center: The knowledge-learning-instruction (KLI) framework: Toward bridging the science-practice chasm to enhance robust student learning. Tech. Rep. CMU-HCII-10-102, Human-Computer Interaction Institute, Carnegie Mellon University (June 2010)
9. Leuski, A., Traum, D.: NPCEditor: A tool for building question-answering characters. In: Proceedings of The Seventh International Conference on Language Resources and Evaluation (LREC) (May 2010)
10. Pon-Barry, H., Schultz, K., Bratt, E.O., Clark, B., Peters, S.: Responding to student uncertainty in spoken tutorial dialogue systems. International Journal of Artificial Intelligence in Education 16, 171–194 (2006)