Choosing When to Interact with Learners

Lei Qu, Ning Wang, and W. Lewis Johnson
Center for Advanced Research in Technology for Education (CARTE), USC / ISI
4676 Admiralty Way, Suite 1001, Marina del Rey, CA, 90292
leiqu@isi.edu, ning@isi.edu, johnson@isi.edu

ABSTRACT
In this paper, we describe a method for pedagogical agents to choose when to interact with learners in interactive learning environments. This method is based on observations of human tutors coaching students in on-line learning tasks. It takes into account the focus of attention of the learner, the learner's current task, and expected time required to perform the task. A Bayesian network model combines evidence from eye gaze and interface actions to infer learner focus of attention. The attention model is combined with a plan recognizer to detect different types of learner difficulties such as confusion and indecision which warrant intervention. We plan to incorporate this capability into a pedagogical agent able to interact with learners in socially appropriate ways.

Categories and Subject Descriptors
I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence – *Intelligent agents*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Training, help and documentation*.

General Terms
Design, Human Factors

Keywords
Pedagogical agents, plan recognition, human-computer collaboration, intelligent assistants, task modeling, Bayesian Network, interface agents, affective interfaces.

1. INTRODUCTION
Animated pedagogical agent technology seeks to improve the effectiveness of intelligent tutoring systems, by enabling them to interact in a natural, more engaging way. However, work to date has focused mainly on improving the output side of the interface, through the inclusion of expressive, lifelike behaviors [1]. The focus of the work described in this paper is on the input side, to enable the agent to track the learner's activities and infer learner state, so it can initiate interactions with the learner at the appropriate time and in an appropriate manner.

Our approach involves monitoring of the learner's activities, both interface actions and focus of eye gaze. It infers the learner's focus of attention using a Bayesian network [2], which allows reasoning under uncertainty with various sources of information. And it combines the method for tracking learner focus of attention with a plan recognition capability for interpreting learner actions and forming expectations of future actions.

This approach should be of general use for improving the interactivity of intelligent tutoring systems. Our particular motivation for conducting this work is to create pedagogical agents that are able to interact with learners in more socially appropriate ways, sensitive to rules of politeness and etiquette and able to influence learner motivational as well as cognitive state [3, 4].

2. BACKGROUND TUTORING STUDIES
In an earlier study, we investigated how human tutors coach learners while interacting with Virtual Factory Teaching Systems (VFTS) [4, 5], an on-line factory system for teaching industrial engineering concepts and skills. We found that tutors used the following types of information, observed and inferred, in deciding when and how to interact with the learner:

- The task that the learner was expected to perform next.
- The learner's focus of attention.
- The learner's self-confidence, inferred from the questions the learner asked.
- The learner's effort expended, as evidenced by the amount of time that the learner spent reading the tutorial and carrying out the tasks described there.

We therefore designed the user interface of our new system to enable an agent to have access to sufficient information about the learner, her/his activities, cognitive and motivational state. The new interface includes three major components:

- The VFTS interface, which reports each keyboard entry and mouse click that the learner performs on it.
- WebTutor, which is an on-line tutorial used to teach learner instruction and concepts of industrial engineering.
- Agent Window, in which the left part of this window is a text window used to communicate with the agent (or a human tutor in Wizard-of-Oz mode) and the right part is an animated character that is able to generate speech and gestures.

The input devices consist of keyboard, mouse, and a small camera focused on the learner's face. This interface thus provides information that is similar to the information that human tutors use in tracking learner activities.

3. OVERVIEW OF OUR METHOD
There are four components in our approach to choosing interaction points:

IUI'04, Jan. 13–16, 2004, Madeira, Funchal, Portugal.
ACM 1-58113-815-6/04/0001.

- WebTutor provides information about what task the learner is working on, as well the actions the learners perform as they read through the tutorial.
- The plan recognizer in VFTS monitors the learner's actions and tracks learner progress through the task.
- The focus of attention module takes input from the WebTutor interface, the VFTS interface and Agent Interface as well as eye gaze information, in order to infer focus of attention.
- Focus of attention and plan recognition information are used to infer learner difficulties such as confusion.

These four parts can provide agents with information about learners' states and their expected tasks. Therefore agents are able to detect learners' confusion, interact with them and help them overcome their difficulties.

Our method is similar to that of the Lumière Project, which also tracks learner activities by monitoring learner actions and tracking learner tasks. The differences are that our system has more information about learner activity (e.g., eye gaze) as well as more information about learner task (from the WebTutor). It can therefore track learner activities with greater confidence, and can therefore focus more on detecting and categorizing learner difficulties.

4. TRACKING LEARNER'S FOCUS OF ATTENTION UNDER UNCERTAINTY

Information about eye gaze is extremely useful for detecting user focus of attention. In our system we want an eye tracking approach that is unobtrusive, that requires no special hardware and no calibration.

We use a program developed by Larry Kite in the Laboratory for Computational and Biological Vision at USC to track eye gaze. It estimates the coordinates on a video display that correspond to the focus of gaze. The agent uses two types of information to infer the learner's focus: (1) information with certainty, i.e., mouse click, type and scroll window events in VFTS, WebTutor and Agent Window, and (2) information with uncertainty, namely data from eye track program and inferences about current state based upon past events. We use a Bayesian network to combine these different sources of information, as shown in Figure 1.

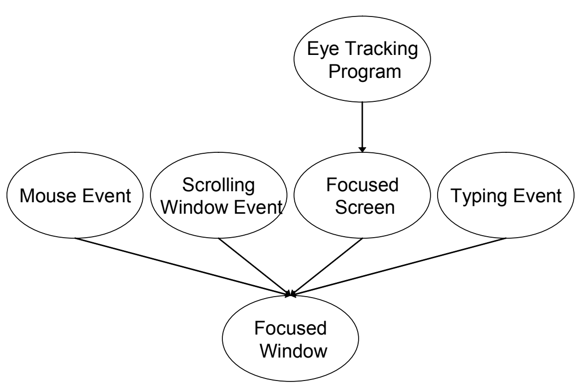

Figure 1: The Bayesian network for inferring focus of attention.

The following variables are employed in this Bayesian network. "Eye Tracking Program" represents where the focus of learner is based on eye tracking program. "Mouse Event" represents where the mouse click event occurs. "Scrolling Window Event" represents whether or not a scroll window event occurred in WebTutor (there is no scroll bar on the VFTS Window). "Focused Screen" represents which part of the screen the learner is focusing on. "Typing Event" indicates where the type events occur.

We ran 12 experiments to explore the performance of the Bayesian network model. In all of these experiments, the agent can get sufficiently accurate data based on the Bayesian network model, and is able to detect periods of fixation comparable to what learners do in practice, i.e., how long learners have already focused on the VFTS or WebTutor.

With the introduction of eye gaze tracking, agent can infer more accurate data for learner's focus of attention and periods of fixation. And the Bayesian network can provide the agent the capability to select from different sources of information to improve the accuracy of the model when inferring learner's focus of attention.

5. PLAN RECOGNITION SYSTEM

To help pedagogical agents determine when to help learners, we need to be able to track the learner's actions as well as track the learner's focus. We created a plan recognition system to help detect possible intervention points. It also serves other functions in determining how to help the learner, which go beyond the scope of this paper.

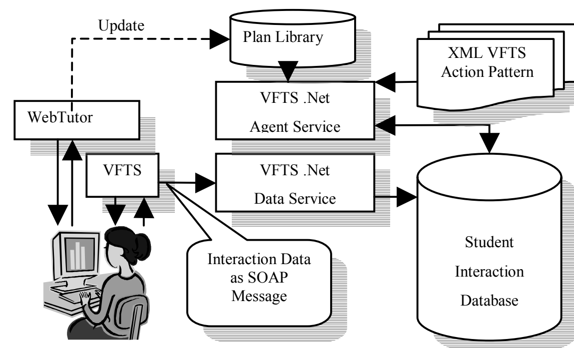

Figure 2: Plan recognition system in VFTS.

The plan recognition system, as shown in Figure 2, has 4 main components: .NET server, Student Interaction Database (SID), Plan Library and Action Pattern file in VFTS.

The .NET server has 2 services: data service and agent service. While students interact with the VFTS client, all interaction data are captured and encoded into SOAP messages and then sent to the data service and saved in SID. Plan recognition is implemented in the .NET agent service. The plan recognizer monitors updates in the SID, retrieves the current plan from the plan library and steps needed to finished a plan from action pattern file, compares user interactions with the current plan, monitors user progress on the current plan and saves progress information in the SID for use by the agent and other tools. A plan in the plan library consists of a set of tasks the user needs to achieve. The preconditions and effects of the steps made up a plan are given in action pattern file.

In the current plan recognition system, we only categorize learner actions which are actions expected by the plan recognizer, and which are actions not expected by the plan recognizer for the current task. In the future, we'd like to further analyze the unexpected actions to find additional cases when the agent might need to intervene, e.g., when the learner is repeatedly performing an inappropriate action.

6. DECIDING WHEN TO INITIATE INTERACTION

The above analyses make it possible for the agent to initiate interactions with the learner so as to maximize positive effect on the learner and minimize negative effects. These include the following:

- Proactively offering assistance when the learner is focusing on a particular task but is failing to make progress on it, and
- Offering assistance when the learner has failed to complete a task and has moved on to another task.

However, a complete solution to the problem of deciding when to intervene with a learner depends upon a number of additional factors:

- The immediate history of past learner performance,
- The learner's individual characteristics (e.g., whether or not they prefer to work on other own),
- Motivational state (e.g., self-confidence),
- Affective state (e.g., is the learner confused or frustrated),
- The degree of disruptiveness of the offered help (e.g., does the agent's comment require an explicit response from the learner), as well as
- The relationship that the agent has established with the learner (e.g., does the learner trust the agent's advice).

Access to this information can permit the agent to be more selective in choosing when to provide feedback, e.g., provide more confirmatory feedback to learners who lack self-confidence. Some of these factors can be derived through further analysis of the learner's activities, as described below.

6.1 Key Parameters Relevant to Interaction

6.1.1 Effort

Effort is an important indicator of intrinsic motivation in learners, and expert human tutors often praise learners for expending effort even when they are not successful [6].

6.1.2 Indecision

Indecision defines the degree of hesitancy to make decisions. In a related fashion, it is possible in some contexts to detect instances of learner *frustration,* as episodes where the learner spends a significant amount of time on a task, tries multiple actions, but fails to make progress in completing the task.

6.1.3 Self-confidence

Self-confidence represents the confidence of learners in the learning environment. If learners perform actions in VFTS after reading tutorial without much hesitancy, such learners must have high confidence. The self-confidence factor can also be reported by the learner via the WebTutor interface.

7. CONCLUSIONS AND NEXT STEPS

In this paper, we present our work on enabling pedagogical agents to track learner activity and focus of attention, to generate analyses that can be used to determine when best to interact with the learner. This work is an important step toward creating interface agents that can interact with learners in a pedagogically effective, socially intelligent fashion.

As part of our future work, evaluations of the adequacy and coverage of this system are planned to commence shortly. When learners study in our environment, we will record the intervention time of the agent and the information of the learners' states and tasks as inferred by the agent. Learners and human tutors can then replay these data in the system and evaluate the performance of agent (e.g., Does the agent make a decision to intervention in appropriate time? Does the inferred focused of attention in agent match the learner's focus? Does plan recognition infer the right goal based upon learner's action?). Furthermore we wish to extend the user monitoring capability to handle a wider range of ambiguous contexts. Based upon these results, pedagogical agent can then interact with learners through a conversational system in more socially appropriate ways.

8. ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant No. 0121330, and in part by a grant from Microsoft Research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

9. REFERENCES

[1] Johnson, W.L., Rickel, J.W., and Lester, J.C. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11, 47-78, 2000.

[2] Pearl, J., Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, CA: organ-Kaufmann, 1998.

[3] W. L. Johnson. Using Agent Technology to Improve the Quality of Web-Based Education. In *N. Zhong and J. Liu (Eds.), Web Intelligence*. Springer, Berlin, 2002.

[4] W. Lewis Johnson. Interaction Tactics for Socially Intelligent Pedagogical Agents. In *Proceedings of the Intelligent User Interfaces*, 2003.

[5] Dessouky, M.M., Verma, S., Bailey, D., & Richel, J. A methodology for developing a Web-based factory simulator for manufacturing education. *IEEE Transactions*, 33, 167-180, 2001.

[6] T. del Soldato and B. Du Boulay. Implementation of motivational tactics in tutoring systems. *Journal of Artificial Intelligence in Education*, 6(4), 337-378, 199.