Representing Adaptive Course Navigation in the Generalized Intelligent Framework for Tutoring

Robert A. Sottilare
US Army Research Laboratory
robert.a.sottilare.civ@mail.mil

Abstract

This paper explores the use of Markov Decision Processes (MDPs) in support of adaptive course navigation in the Generalized Intelligent Framework for Tutoring (GIFT). GIFT is an open source architecture for authoring and evaluating Intelligent Tutoring Systems (ITSs) and adaptive course navigation is an AI-based technique which considers attributes of the learner and the instructional context to select actions which will optimize learning. GIFT's current adaptive course navigation model is decision tree-based. Other ITSs primarily use performance as a driver for navigation without consideration for other learner states. The adaptive course navigation model presented aligns closely with the principles of MDPs where a user's current state, possible actions and a reward function determine movement to a future state. Unlike decision trees used which are currently used in GIFT, MDPs also account for multiple states to determine future states and also consider uncertainty in the assessment of learner states.

Introduction

Sottilare (2012) developed an adaptive learning effect model (LEM) to represent optimal interaction between the learner, tutor, and the instructional environment (Figure 1).

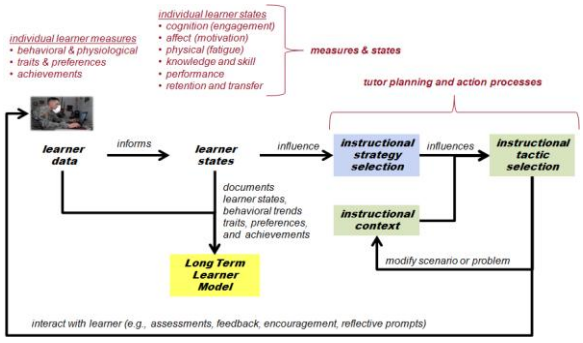

Figure 1. Learning Effect Model (LEM) for Individuals.

This model has been elaborated over time to account for real-time interaction for both individuals and teams (Fletcher & Sottilare, 2013; Sottilare, 2015), assessment of learner competencies across a variety of training tasks involving learning progressions (e.g., marksmanship), and assessment of skill development across long time spans which might include career-wide progressions.

Similar to Vygotsky's Zone of Proximal Development (ZPD; 1978) and tutoring design principles developed by Anderson, Boyle, Farrell and Reiser's (1987) and later elaborated by Corbett, Koedinger and Anderson (1997), Sottilare and Goldberg (2012) envisioned that cognitive load (e.g., working memory) could be optimized by matching the domain competence of the learner and the difficulty/grain size of the problem presented to the learner. In other words, provide more difficult and elaborate problems to highly skilled learners and easier, straight-forward problems to less skilled learners.

From a practical standpoint, three areas of learner assessment are critical in order to provide optimal instruction and expose the learner only to the concepts required for new learning. First, the ITS must understand the level of the learner's prior domain knowledge. Second, it must be able to assess when the learner has mastered new knowledge so they can proceed to new concept or learning objectives. Third, in order to keep the learner motivated and focused on learning, the ITS must understand when new instructional content elicits emotion in the learner which might either enhance (e.g., joy, dominance) or detract (e.g., boredom, frustration, long term confusion) from learning.

The LEM and its associated instructional theory (e.g., cognitive load theory, and component display theory) are central to how GIFT guides instruction. GIFT is a largely domain-independent architecture which is focuses on reuse and best practices to reduce the time and skill needed to author complex ITSs. ITSs developed with the GIFT authoring tools also have embedded instructional theory to optimize the development of learner knowledge and skills.

Component Display Theory (CDT; Merrill, Reiser, Ranney, and Trafton, 1992) arranges instruction into four phases or quadrants to insure the learner: understands basic domain terms and principles (rules quadrant); is exposed to domain-relevant models of success (examples quadrant); can remember information from the rules and examples quadrants (recall quadrant); can successfully apply knowledge to structured application to build skills (practice quadrant).

In addition to CDT, GIFT also uses a decision tree structure to align learner attributes with recommended instructional strategies to optimize learning. This decision tree considers learner attributes such as prior knowledge and motivation/interest to align instructional content. The decision tree is based on a large review and meta-analysis of the instructional literature. For example, the pedagogical request in GIFT for a motivated journeyman in the rules quadrant is for content which is "text/visual information, of moderate difficulty, and of interactive multimedia instructional level 2". The primary drawback to the decision tree structure is that it assumes learner attributes with 100 percent certainty. For example, instructional recommendations based on a learner state of confusion could produce negative learning experiences if it turned out that the learner's actual state was boredom or frustration. A method is needed to deal with uncertainty related to learner state and this is our primary motivation in exploring the use of MDPs within GIFT.

Decision Tree Course Navigation in GIFT

As noted in the LEM (Figure 1), the real-time interaction between the learner and the instructional environment provide the basis for decision-making by the ITS. The tutor's knowledge of the learner is central to optimizing this decision-making. The tutor assesses the learner's prior domain knowledge, analyzes their physical and behavior cues to assess the learner's state, and then uses this information to select strategies and ultimately apply tactics (actions) that can either affect the learner directly (e.g., support, hints, prompts, questions) or affect the complexity of the instructional environment and thereby the learner indirectly.

The part of the LEM that takes action is the domain model. The domain model also assesses the learner's progress toward learning objectives or concepts as they are referred to in the GIFT ontology (Sottilare, 2012). As the tutor decides what to do at the next turn, it considers the learner's states and traits to plan a strategy. It also considers the instructional context (where the learner is in the course) and the recommended strategy (e.g., prompt the learner for additional information) to select a tactic (select and appropriate question and ask it). Next we discuss how tactical decisions are made in GIFT and how those interactions guide the learner through a lesson.

Decision Tree-Based Course Navigation in GIFT

Before exploring a new adaptive method for course navigation in GIFT, we will review how GIFT currently guides learners through a lesson. For this example (Figure 2), we assume a sequential lesson with three concepts (A, B, and C) which must be mastered by the learner.

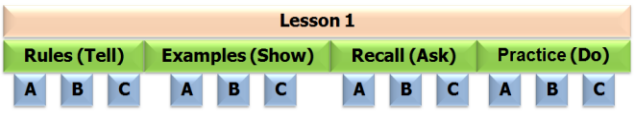

Figure 2. Sequencing of Concepts (A, B and C) in a GIFT Lesson

As noted previously, GIFT uses CDT to adapt course flow for each learner. In our example, the learner must master concepts A, B, and C and each is presented to the learner sequentially in each of the CDT quadrants. For the rules quadrant, the content needed to illustrate the basic terms and principles of concepts A, B, and C are presented to the learner. Next examples of n are presented for all three concepts. After that the learner is assessed on their ability to recall the information provided in both rules and examples. Finally, the learner is asked to apply their knowledge in a practice environment to develop skills and show progress toward objectives (mastery of concepts).

In our decision tree architecture, GIFT examines a tuple composed of a set of learner states (e.g. domain knowledge, cognitive load, and/or affective), an environmental state (where the learner is in the context of the lesson), a set of actions available to the tutor based on the environmental state, and a set of rewards based on progress toward mastery (performance state). The ITS may use all of these attributes to drive its pedagogical decisions in an effort to learn all three concepts in the least time possible. With the exception of the learner's states, all are known with reasonable certainty.

According to Sottilare, Ragusa, Hoffman & Goldberg (2013), the LEM's learner data and states may be further decomposed and used by the tutor to select optimal strategies and tactics. The learner data may include values, preferences, interests, goals, behaviors, and physiological measures which can be used directly by the tutor to make strategic/tactical decisions or can be used to interpret learner states. Sottilare, et al (2013) also note that learner states may include: potential (based on prior domain knowledge); performance; cognitive load; affective states (e.g., personality or emotions); motivational state which is influenced by goals, preferences, and interests; and physical state (e.g., fatigue, level of motor skills). Motor skills

may be measured in terms of speed, precision, distance, or adherence to a particular set of procedures or techniques.

In Figure 3, we have a learner who is highly motivated to learn about a particular domain and the learner's knowledge of the domain is moderate (journeyman). For each CDT quadrant, GIFT provides a recommended strategy related to the type of content (e.g., text, visual, case study) to be presented, the complexity or difficulty, and the interactive multimedia instruction (IMI) level. These recommendations are based on a large scale meta-analysis of the training and education literature.

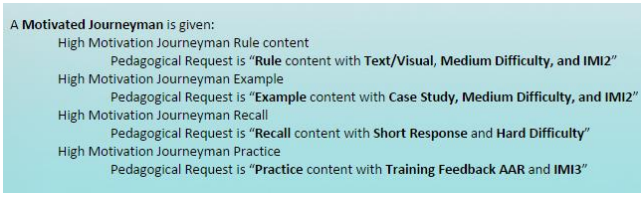

Figure 3. Examples of Tutor Decisions in GIFT

Agent-Based Approach to Course Navigation

According to Mitchell (1997), an agent-based approach to reinforcement learning is focused on the goal of learning to choose actions that maximize current and future rewards. The desirable characteristics of these agents are: reactive, proactive, and cooperative. Reactive agents should be responsive to change(s) in the environment and active in enforcing policies or rules. Proactive agents should take the initiative to focus on achievement of long-term goals, recognize opportunities, and learn and adapt to optimize learning. Cooperative agents should share information and act together to achieve long-term goals.

For our purposes, in our MDP, the agent should be able to recognize a set of distinct states (S) from which a set of finite actions (A) can be performed. Individual actions (a) should result in movement to a next state (s) and an associated reward (r). At each turn, the tutor will assess the current state of the learner and their performance, and provide a reward based on progress in mastering concepts.

For an agent-based model, Figure 4 shows the relationship between states, actions, and rewards for the learner, the environment (instructional content) and the agent. The agent monitors the state of both the learner and the environment (instructional content) and can change the level of interaction with the learner (e.g., support) or the level of difficulty of the environment.

Rewards can change based on progress toward goals or concepts or by demonstrating higher skills by solving more difficult problems in the environment. Interaction between the learner and the environment is typical of non-adaptive training systems where the learner can observe the environment and act on it, but neither the learner nor the environment can determine rewards. The cumulative value (V^{Π}) achieved by following a policy (Π) from an initial state (s_t) follows:

$$V^{\Pi}(s_t) \equiv r_t + \gamma\, r_{t+1} + \gamma^2\, r_{t+2} + \ldots \equiv \sum_{i=0}^{\infty} \gamma^i\, r_{t+i}$$

where γ ($0 \leq \gamma < 1$) is a constant that determines the relative value of rewards. For $\gamma = 0$, only the immediate reward is considered for all states and as γ approaches 1, future rewards are given greater and greater consideration over rewards associated with initial states.

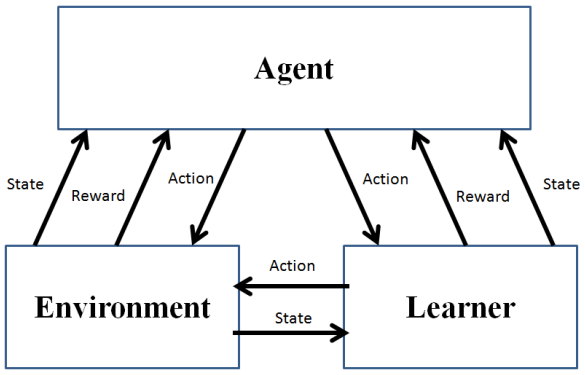

Figure 4. States, Actions, and Rewards

Figure 5 illustrates positive (green arrows) and negative (red arrows) rewards for actions in the environment (specifically in the rules quadrant). In this model the learner is rewarded for taking actions to acquire new knowledge, but not rewarded for going back to review old knowledge.

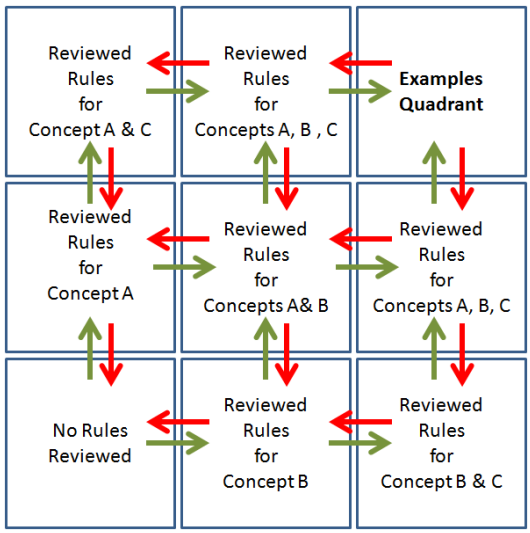

Figure 5. Learner Performance States (progress toward objectives where each box represents a learner's performance state)

The same type of reward system could easily be implemented in GIFT for the examples, recall, and practice quadrants of CDT. Probabilities for a small number of possible actions available in each state could be ascertained over time and the negative assessments in recall and practice could form a reinforcement learning policy similar to the Q function (Mitchell, 1997, p. 374-377). For example, the agent would observe the current state (*s*), and then select an action (*a*) based on learner actions. The ITS would then receive an immediate reward and observe a new state (*s'*). By recording and comparing/contrasting rewards over time, this method allows the tutor to learn how to select highest rewards states and how to avoid incorrect learner state assessments.

Discussion

In this paper, we have contended that the reinforcement learning strategies associated with MDP offer distinct advantages over rule-based or decision-tree based AI solutions. In scenarios where learner states can vary widely and are often difficult to determine, MDPs offer more flexibility in acquiring and analyzing learner data to determine states and matches to optimal outcomes. MDPs also allow for a variety of outcomes/rewards given the same state and action whereas decision trees and rule-based solutions offer only a single path. For instance, we might select a different, but optimal instructional tactic for identical learner states and environmental conditions because the values or preferences in individual learners are different.

A central question to optimizing the selection of strategies and tactics is what desired outcome or outcomes are related to the reward state in our MDP. In our LEM, effect size is a measure for quantifying the difference between the rewards resulting from multiple strategic or tactical options. Learning effect is a measure of the difference between instructional methods on learning gains. For our purposes we have selected four important outcomes: accelerated learning, deep learning and enhanced retention, enhanced performance, and enhanced transfer as discussed below.

Accelerated learning results when adaptive instructional methods decrease the amount of time needed to acquire a unit of knowledge or develop a unit of skill compared to traditional (currently implemented) instructional methods. It may be possible to accelerated learning by skipping old instructional content based on an individual's domain competency. It may also be possible to accelerate learning by understanding when a learner has achieved mastery of a concept and then moving them to the next concept as soon as practical.

Retention is the ability to maintain a level of knowledge and skill to remain proficient in a particular task. Deep learning and enhanced retention can result when learners encounter "desirable difficulties" (Bereiter & Scardamalia, 1985; Bjork, 1988) and are challenged to work harder during initial learning experiences. This invites "deeper processing of material than people would normally engage in without explicit instruction to do so" (Bjork, 1994). The idea that desirable difficulties can gel learning and support longer term retention is a principle adopted within GIFT's pedagogical structure in the form of "indirectness" as defined in the INSPIRE model of tutoring (Lepper, Drake, and O'Donnell-Johnson, 1997). Desirable difficulties are an important adaptive tutoring strategy and they are closely related to Vygotsky's ZPD (1978), but instead of matching the competence level of the learner and the instructional material, we are challenging the learner to reach.

Most structured practice is geared toward enhancing performance. Whereas learning is the acquisition of knowledge and skills, performance is the result of applying knowledge and skill. There is often a divide between practitioners (trainers) and theorists (educators) with respect to performance. The focus for practitioners is to enhance performance to varying levels of automaticity without the need to understand why certain methods are used. Theorists tend to focus not only on what is being done, but why.

Finally, the transfer of skills from one environment (e.g., training system) to another environment (e.g., operational system) is an important element of learning and a consideration for an agent-based system concerned with optimal instruction. The selection of instructional methods may need to vary based on the weight of the desired outcome. For instance, if transfer is more heavily weighted than near term performance for a learner with significant domain experience, we might want to place more emphasis on instructional tactics that align more closely with how the task is performed in the operational environment to promote higher transfer of skills.

A drawback to any adaptive solution is that the ITS author would then need to identify each and every one of the learner state transitions that should result in the delivery of an instructional strategy (e.g., if the learner is frustrated, then the author might adapt the scenario to be easier). In some instances, a defined state transition may happen more than once, in which case, the author would need to provide more than one instructional strategy to choose from and this is an additional authoring burden. Furthermore, each strategy then has one or more tactics (an action based on a plan or strategy) to choose from. While this flexibility allows GIFT to move away from a fixed state diagram and introduce probabilities, it comes at a cost to the author.

The reward structure examined in this paper was limited to optimizing near term reward (discounting) at the expense of life-long learning. There are many instances where near term performance could and should be sacrificed for the sake of learning valuable lessons which allow

the learner to more easily retain and transfer knowledge and skills to a broader array of domains. In the same instance, we could also see the possibility of sacrificing learning to enhance the confidence or esprit de corps of low performing learners.

MDP Applied to GIFT

As noted, once GIFT makes a selection based upon an imprecise assessment of a learner's state, there is no mechanism to validate the accuracy of that learner state in the future or to adjust the decision tree to provide another recommendation based on identical conditions. MDPs offer the capability to examine options for moving forward to the next performance state and enable reinforcement learning over the long term to support continuous improvement of strategy and tactics selections.

States and Rewards

So, how should the state, *s*, be represented in an agent-based GIFT? Based on the models shown in Figures 4 and 5, elements of the learner's state and the environment are necessary to select optimal strategies and tactics within GIFT's architecture. For the learner's state, we need to represent the following elements: prior domain knowledge; concepts under instruction; progress toward learning objectives at any time, *t*, in the instructional process; emotional states that are moderators of learning. In defining the components of a state, s, we have noted the importance of four sub-states. These sub-states provide a mix of learner states and a list of potential actions by the tutor where maintaining positive learner states and progress toward learning objectives are associated with higher rewards. The following is a decomposition of the elements of those sub-states:

- *Prior domain knowledge* (competence level = novice, journeyman, expert)
- *Concepts under instruction* (list of concepts = A, B, C; order of instruction)
- *Progress toward learning objectives* (rules reviewed for A, B, C; examples reviewed for A, B, C; recall tested for A, B, C; skills tested for A, B, C)
- *Emotional states* (states observed; negative states moderated; positive states maintained)

Optimizing Actions in GIFT

Movement from one state to another results in a positive reward when a concept is either reviewed (rules quadrant), reviewed (examples quadrant), assessed (recall quadrant), or assessed (practice quadrant). Additional rewards are provided when moving from one quadrant to another, and when demonstrating mastery of a concept within a quadrant.

Below are recommended implementations for an agent-based course navigator for GIFT based on our four desired outcomes: accelerated learning, deep learning and enhanced retention, enhanced performance, and enhanced transfer.

For Q learning (Mitchell, 1997), the agent is attempting to learn an optimal policy (Π^*) where the optimal action (a) in state (s) is the action that maximizes the sum of the immediate reward (r) plus the value of the immediate successor state as discounted by γ:

$$\Pi^* (s) \equiv \text{argmax} [r (s, a) + \gamma V^* \delta (s, a)]$$

To optimize actions for our accelerated learning outcome, the immediate reward is most important as it lessens decisions to review old material and thereby accelerates learning (see Figure 5). Therefore, $\gamma = 0$ since it reduces the value of the immediate successor state to 0 and rules out any rewards with negative values (e.g., old information).

To optimize actions for our deep learning and enhanced retention outcome, the immediate reward is less important as it increases decisions to review old material and thereby deepens learning (see Figure 5). Therefore, as γ approaches 1, it increases the value of the immediate successor state over the immediate reward.

To optimize actions for our enhanced performance outcome, the immediate reward is important, but so is overall value. Therefore, as γ should be adjusted to optimize the sum of value of the immediate reward and the immediate successor state.

Finally, to optimize actions for our enhanced transfer outcome, the immediate reward is less important than later rewards. Therefore, as γ approaches 1, it increases the value of the immediate successor state over the immediate reward thereby optimizing actions which enable higher transfer from the training environment to the operational environment.

Conclusions

We provided a potential solution for expanding the flexibility and introducing stochastic elements into the adaptive instructional process. This could be done without major changes to the GIFT architecture, but comes with a cost in

terms of time and skill in the authoring process. Since each action in the MDP must be accounted for in terms of actions, the MDP solution should also be paired with AI-based solutions for simplifying the authoring process for ITSs.

MDPs from a stochastic point of view are also attractive alternatives to decision-trees based on the ability to project expected value or expected total reward into the future based on known decision chains.

This paper examined elements of the MDP related to the instruction of individuals. The development of MDPs for team-based activities poses a much more significant challenge in terms of complexity. State assessments for teams are more complex and less accurate so generalized rules based on the team instruction literature applied to the initial states in an MDP could lead to some experimental strategies which yield future best practices.

References

Anderson, J., Boyle, C., Farrell, R., & Reiser, B. (1987). Cognitive principles in the design of computer tutors. In P. Morris (Ed.), *Modeling cognition*. NY: John Wiley.

Bereiter, C., & Scardamalia, M. (1985). Cognitive coping strategies and the problem of "inert knowledge". In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Vol. 2. Current research and open questions* (pp. 65-80). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (1988). Retrieval practice and maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.). Practical aspects of memory: *Current research and issues. (Vol 1),* pp. 396-401. NY: Wiley.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185-205). Cambridge, MA: MIT Press.

Corbett A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (pp. 849–874). Amsterdam: Elsevier.

Goldberg, B., Brawner, K., Sottilare, R, Tarr, R., Billings, D. & Malone, N. (2012) Use of Evidence-based Strategies to Enhance the Extensibility of Adaptive Tutoring Technologies. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Arlington, VA : National Training Systems Association.

Lepper, M. R., Drake, M., & O'Donnell-Johnson, T. M. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds), *Scaffolding student learning: Instructional approaches and issues* (pp. 108-144). New York: Brookline Books.

Merrill, D., Reiser, B, Ranney, M., and Trafton, J. (1992). Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. *The Journal of the Learning Sciences*, 2(3), 277-305.

Mitchell, T. (1997). Machine Learning. WCB/McGraw-Hill.

Sottilare, R. and Goldberg, B. Designing Adaptive Computer-Based Tutors to Accelerate Learning and Facilitate Retention. *Journal of Cognitive Technology: Contributions of Cognitive Technology to Accelerated Learning and Expertise* 2012, 17, 1, 19–34.

Sottilare, R. (2012). Considerations in the development of an ontology for a Generalized Intelligent Framework for Tutoring. *International Defense & Homeland Security Simulation Workshop in Proceedings of the I3M Conference*. Vienna, Austria, September 2012.

Sottilare, R., Ragusa, C., Hoffman, M. & Goldberg, B. (2013). Characterizing an adaptive tutoring learning effect chain for individual and team tutoring. In Proceedings of the *Interservice/Industry Training Simulation & Education Conference*, Orlando, Florida, December 2013.

Sottilare, R., (2015). Challenges in Moving Adaptive Training & Education from State-of-Art to State-of-Practice. In Proceedings of the "Developing a Generalized Intelligent Framework for Tutoring (GIFT): Informing Design through a Community of Practice" Workshop at the *17th International Conference on Artificial Intelligence in Education (AIED 2015)*, Madrid, Spain, June 2015.

Vygotsky, L. S. (1978). Mind in society--The development of higher psychological processes. Cambridge: Harvard University Press.