

Generalizability of Goal Recognition Models in Narrative-Centered Learning Environments

Alok Baikadi, Jonathan Rowe, Bradford Mott, and James Lester

North Carolina State University, Raleigh, North Carolina, USA
{abaikad, jprowe, bwmott, lester}@ncsu.edu

Abstract. Recent years have seen growing interest in automated goal recognition. In user-adaptive systems, goal recognition is the problem of recognizing a user's goals by observing the actions the user performs. Models of goal recognition can support student learning in intelligent tutoring systems, enhance communication efficiency in dialogue systems, or dynamically adapt software to users' interests. In this paper, we describe an approach to goal recognition that leverages Markov Logic Networks (MLNs)—a machine learning framework that combines probabilistic inference with first-order logical reasoning—to encode relations between problem-solving goals and *discovery events*, domain-specific representations of user progress in narrative-centered learning environments. We investigate the impact of discovery event representations on goal recognition accuracy and efficiency. We also investigate the generalizability of discovery event-based goal recognition models across two corpora from students interacting with two distinct narrative-centered learning environments. Empirical results indicate that discovery event-based models outperform previous state-of-the-art approaches on both corpora.

Keywords: Goal recognition, narrative-centered learning environments, intent recognition.

1 Introduction

A key challenge in user-adaptive systems is recognizing users' intentions and plans. *Goal recognition*, also known as intent recognition, is the process of identifying the high-level objective that a user is pursuing based on an observed sequence of actions performed by the user. It is a restricted form of plan recognition, which involves identifying both the user's goal and the action sequence that will achieve the goal. The capacity to reason about users' plans, activities, and intentions enables user-adaptive systems to predict actions that will be performed by the user. Computational models of goal recognition have been investigated across a range of applications, including intelligent tutoring systems [1], dialogue systems [2], and digital games [3, 4].

One type of user-adaptive system that particularly stands to benefit from goal recognition models is narrative-centered learning environments. Narrative-centered learning environments combine the personalized educational support of intelligent tutoring systems with the engaging, interactive storylines of digital games [5, 6]. In narrative-centered learning environments, goal recognition models can be used to

support struggling students by detecting erroneous problem-solving approaches or misconceived strategies that require remediation. Models for goal recognition can also be used to pre-emptively adapt learning environments’ interactive narratives, dynamically adjusting difficulty or providing personalized scaffolding and guidance.

In this work, we investigate the generalizability of goal recognition models that leverage Markov logic networks (MLNs) [7] to encode relations between students’ goals and problem-solving progress in narrative-centered learning environments. Markov logic is a statistical relational learning framework that combines probabilistic inference with first-order logical reasoning. In our work, we examine the impact of a domain-specific representation of problem-solving progress, called *discovery events*, on MLN-based goal recognition models’ accuracy and efficiency. Discovery events encode major plot milestones that convey key problem-solving information to students during narrative-centered learning. We examine the impact of incorporating logical representations of discovery events into MLN-based goal recognition models, comparing discovery event-based models to a MLN-based baseline approach.

To investigate the generalizability of the models, we train and evaluate them using corpora from two distinct narrative-centered learning environments: a narrative-centered learning environment for middle school microbiology that centers on diagnostic problem solving and another narrative-centered learning environment for elementary science education that emphasizes exploration-focused quests on maps, models, and landforms. Empirical results from the evaluation indicate that discovery event-based goal recognition models outperform prior state-of-the-art MLN approaches. Furthermore, the findings generalize across both learning environments, indicating that discovery event-based MLN models for goal recognition are an effective approach for predicting students’ goals during narrative-centered learning.

2 Related Work

Goal recognition, along with the related tasks of activity recognition and plan recognition, has been the subject of considerable interest in the AI community for decades. A wide variety of computational frameworks, including symbolic representations [8–10], probabilistic representations [11, 12], grammar-based models [12, 13], and hybrid models [2, 14, 15] have been investigated extensively, both from theoretical and empirical perspectives. Goal recognition models have also been examined across a range of applications, including dialogue management [16], learning environments [1], cybersecurity [10], story understanding [11], and games [3, 4, 15, 17, 18].

Over the past few years, Markov logic networks have received growing attention for their promise in plan, activity, and goal recognition [15, 19, 20]. Sadilek and Kautz [15] used Markov logic to investigate activity recognition in a multi-agent Capture the Flag game using GPS data. In their model, a Markov logic network combines hard and soft constraints derived from Capture the Flag rules to de-noise and label GPS data in terms of “capture” events. Experiments demonstrated that the MLN model significantly outperformed alternate probabilistic and nonprobabilistic approaches, correctly identifying 90% of capture events. Singla and Mooney [19] devised a meth-

od for constructing MLN-based plan recognition models using abductive reasoning over planning domains. Experiments found that Singla and Mooney’s approach improved predictive accuracy over competing techniques. However, by framing the problem in terms of abductive inference, their approach requires a formal description of the planning domain be available.

We extend prior work on goal recognition in narrative-centered learning environments by examining the impact of logical representations of discovery events on the accuracy and efficiency of goal recognition [20, 21]. Further, we investigate discovery event-based models’ capacity to generalize across multiple narrative-centered learning environments.

3 MLN-Based Goal Recognition with Discovery Events

We formalize goal recognition as a classification task: given a sequence of actions performed in a narrative-centered learning environment, determine the goal that is most likely associated with the actions. We use Markov logic networks [7], a statistical relational learning framework, to perform classification. *Markov logic networks* (MLNs) unify logical and probabilistic representations for reasoning over a knowledge base. A MLN is a set of first-order logical formulae with associated weights. Each weight denotes the formula’s strength as a constraint on the set of possible worlds described by the knowledge base. During inference, the formulae and weights are transformed into a Markov network, a graphical model with each node representing a random variable, and each undirected edge representing conditional dependencies. The joint probability distribution is defined as $P(X = x) = \frac{1}{Z} \prod_k \phi_k(x_{\{k\}})$, where ϕ_k is a potential function over a clique in the graph. The potential function is derived from the weight of a formula and the groundings of the individual predicates in the training corpus.

For this work, weights are learned using Markov: The Beast, an open source implementation of Cutting Plane Inference for the MLN learning framework [22]. Cutting Plane Inference uses Weighted MAX-SAT [7] inference by incrementally adding violated formulae to a knowledge base and solving the remaining Markov network.

Classification within MLNs is similar to other probabilistic graphical models. A random variable, or predicate in MLNs, is created for the class to be predicted. During inference, probabilities for different groundings of the hidden predicate are computed for each time step in the student interaction data.

In order to prepare our goal recognition corpora for training, validating, and evaluating classification models, we make several assumptions about our domains. First, we assume students only pursue a single goal at a time. Second, we assume that the distribution between observations and goals is fixed across players. Third, we assume that the student is not aware that the system is monitoring their goals, and therefore the system does not influence students’ goal-directed behavior.

Table 1: Predicates for goal recognition

Predicate	Semantics
$action(t, a)$	Observed player performing action a at time t
$argument(t, a)$	Object a was involved in the player's action at time t
$location(t, l)$	Observed player at location l at time t
$state(t, s)$	Player has passed the milestones for state s before time t
$goal(t, g)$	Player is attempting goal g at time t

Based upon these assumptions, we annotate each student action in a goal recognition corpus using the following process. First, the students' interaction traces are scanned for points in which goals are achieved. These observations are annotated with the goal. Because we assume there are no interleaved goals, every remaining action is annotated with the next goal observed. Afterward, the original goal-achieving actions are removed, because it would be trivial to recognize a goal from a goal-achieving action.

Table 1 shows the four predicates used to describe each observation in a narrative-centered learning environment. To represent location symbolically, the virtual environment is discretized, and each region is assigned a unique identifier. These discrete regions are specified as arguments to location predicates. Logical formulae were developed to connect these observed predicates to a hidden predicate representing the student's goal associated with that action. Formulae were included to ensure that only one goal was associated for any given time step.

In addition to formulae connecting the observed predicates to the goal predicates, we introduce new formulae to capture significant milestones in the interaction history: discovery events. Within a narrative-centered learning environment, the overall task can be conceptualized in terms of resolving a central driving question about the plot. Central questions often revolve around the nature of the setting (e.g., Where are we?), what characters will do or have done (e.g., Who committed the crime?), or the consequences of a major event (e.g., What is going to happen now?).

Discovery events represent the user's narrative progress by encoding partial answers to the plot's central driving question. For example, in the context of a mystery plot, discovery events represent clues that the player obtains. To illustrate, consider the following medical mystery scenario. The protagonist, a medical detective, is investigating a disease afflicting a group of sick patients. In order to learn more about the illness, the protagonist asks a sick patient about his symptoms and recent medical history. The resulting conversation is a discovery event; the protagonist gains valuable information to diagnose the illness and resolve the story's central question: "What is making the patients sick?" Later in the investigation, the protagonist runs a laboratory test on some eggs that the sick patient recently consumed. The eggs test positive for salmonella. This is another discovery event that reveals the likely transmission source for the disease. In addition to plot revelations, discovery events can include first-time demonstrations of game mechanics that are required to advance the interactive narrative's plot. For example, when the protagonist demonstrates that she can successfully use the laboratory's testing equipment for the first time, the action is

considered a discovery event. In this manner, discovery events encompass a broad range of events that transpire in interactive narratives.

4 Testbed Narrative-Centered Learning Environments

In order to investigate the effectiveness and generalizability of discovery event-based models of goal recognition, we train and evaluate the models on goal recognition corpora from two distinct narrative-centered learning environments. The first environment, *CRYSTAL ISLAND: OUTBREAK*, is an educational interactive narrative for middle school science that focuses on diagnostic problem solving. The second environment, *CRYSTAL ISLAND: UNCHARTED DISCOVERY*, is an educational interactive narrative for upper elementary science that emphasizes quests about maps, models, and landforms. In this section, we describe each of the learning environments.

4.1 CRYSTAL ISLAND: OUTBREAK

CRYSTAL ISLAND: OUTBREAK is a narrative-centered learning environment for middle school microbiology. The environment features a science mystery where students attempt to discover the identity and source of an infectious disease that is plaguing a research team stationed on a remote island. Students explore the research camp from a first-person viewpoint and manipulate virtual objects, converse with characters, and use lab equipment and other resources to solve the mystery.

In order to represent students' problem-solving progress in *OUTBREAK*'s interactive narrative, we identified seven narrative sub-goals that are central to the solving the science mystery. These goals are never directly presented to the student; rather, they are discovered by the student during the course of gameplay. Five of the goals involve interviewing virtual characters. One goal involves successfully running a laboratory test on the disease's transmission source. The final goal involves submitting a correct diagnosis to a virtual character, thereby solving the mystery.

There are nineteen different types of actions in the *OUTBREAK* environment. These actions include interactions with virtual objects, talking with virtual characters, gathering and recording information in a diagnosis worksheet, and interacting with the testing equipment in a virtual laboratory. There are seven major regions of the island. Each of these locations is further divided into several sub-areas in a rough grid. In total, there are 39 distinct, non-overlapping locations that are recorded throughout interactions with the narrative-centered learning environment. Narrative state is conceptualized in terms of three phases of the plot: an introductory phase, a laboratory testing phase, and a diagnosis worksheet phase. After the introduction, the latter two phases can be attempted in either order. Narrative state is encoded as a binary vector indicating which of the phases have been completed.

In addition to the narrative state, six discovery events were identified in the plot. The discovery events represent moments in the interaction where critical plot-related information is communicated to the student. Three of the discovery events occur during dialogue with virtual characters, two occur during interactions with in-game

books and posters, and two occur during interactions with a diagnosis worksheet tool that students' use in the environment.

4.2 CRYSTAL ISLAND: UNCHARTED DISCOVERY

CRYSTAL ISLAND: UNCHARTED DISCOVERY takes place on a fictional island in the Oceania region of the Pacific Ocean. Students take on the role of a child who has been shipwrecked on an island after a tropical storm along with a crew of explorers. After an introductory cinematic, the student joins a cast of virtual characters on the island in establishing a new life. From there, the student explores a rich 3D environment as she performs tasks for the various island inhabitants. As students learn about landforms, navigation and modeling, they are asked to perform several quests that assess their skills. Following a brief tutorial segment to familiarize students with the interface and gameplay, students have access to a virtual tablet device that provides access to several in-game applications to support problem solving.

In the dataset used in this work, students completed four quests. Two of the quests focused on understanding landforms, such as plateaus, deltas and waterfalls. Two of the quests involved understanding navigation, both through reading a map and following a heading for a specified amount of distance. Each quest was associated with three goals. These goals were used as the targets for goal recognition. Once a quest was begun, its three goals could be pursued in any order. Quests could also be repeated, since faster times were given in-game rewards in the form of trophies.

There were thirty-seven distinct actions that the student could perform in UNCHARTED DISCOVERY. These range from general actions (e.g., moving between areas, managing items), to quest management (e.g., accepting quests, quitting quests), to quest-specific actions (e.g., taking photographs, collecting flags). There are sixty outdoor locations on the island, as well as three buildings that may be entered. The phases of the narrative state are partitioned in terms of the quests; the completion of a quest marks the transition from one phase to another. Because the plot of UNCHARTED DISCOVERY is organized in terms of quests, discovery events are also defined in terms of quests. Two of the discovery events involve quest management, accepting a quest and quitting a quest. Another three occur during interactions with the table. The students are able to view a map of the island, access a virtual encyclopedia, and take freeform notes to organize their thoughts. Finally, three discovery events are quest-specific actions, such as placing signs at specific locations, taking photographs, and picking up colored flags.

5 Goal Recognition Corpora

The OUTBREAK corpus was gathered from 153 8th grade students, aged 12-15 (M=13.3) in a North Carolina public school. Sixteen players were removed due to incomplete data or prior experience with the narrative-centered learning environment. Of the remaining 137 players, 77 were male and 60 were female. A minority, 41.6% of students, completed the mystery in the allotted time. Twenty of the 137 players

$$\begin{aligned}
\forall t : |\forall g : goal(t, g)| = 1 & \quad (1) \\
\forall t, g : goal(t, g) * w_2(g) & \quad (2) \\
\forall t, a, g : action(t, a) \Rightarrow goal(t, g) * w_3(a, g) & \quad (3) \\
\forall t, l, g : location(t, l) \Rightarrow goal(t, g) * w_4(l, g) & \quad (4) \\
\forall t, s, g : state(t, s) \Rightarrow goal(t, g) * w_5(s, g) & \quad (5) \\
\forall t, a, s, g : action(t, a) \wedge state(t, s) \Rightarrow goal(t, g) * w_6(a, s, g) & \quad (6) \\
\forall t, a, g : action(t-1, a) \Rightarrow goal(t, g) * w_7(a, g) & \quad (7) \\
\forall t, l, g : location(t-1, l) \Rightarrow goal(t, g) * w_8(l, g) & \quad (8) \\
\forall t, s, g : state(t-1, s) \Rightarrow goal(t, g) * w_9(s, g) & \quad (9) \\
\forall t, a, s, g : action(t-1, a) \wedge state(t-1, s) \Rightarrow goal(t, g) * w_{10}(a, s, g) & \quad (10) \\
\forall t, a_1, a_2, g : action(t-1, a_1) \wedge action(t, a_2) \Rightarrow goal(t, g) * w_{11}(a_1, a_2, g) & \quad (11) \\
\forall t, a_1, a_2, g_1, g_2 : \left(\begin{array}{l} action(t-1, a_1) \wedge action(t, a_2) \\ \Rightarrow [goal(t-1, g_1) \Rightarrow goal(t, g_2)] \end{array} \right) * w_{12}(a_1, a_2, g_1, g_2) & \quad (12)
\end{aligned}$$

Fig. 1. Baseline formulae

experienced a game crash, and were instructed to restart the game. In these cases, both sets of observations will be used for this work [5].

The students were first presented with an overview of OUTBREAK, which introduced the backstory and task description and the game controls. They were also given handouts, which contained that information, as well as a map of the island and a description of the characters in the virtual environment. The students were then given a pre-survey, which assessed their science content knowledge, in addition to various psychological instruments. The students interacted with the system for a maximum of 60 minutes. At the end of 60 minutes, or when they completed the mystery, the students were given a post-survey. The whole interaction lasted about 120 minutes.

The UNCHARTED DISCOVERY corpus was collected from eight North Carolina public elementary schools. Each fifth-grade classroom interacted with the software over a 4-week period. A total of 831 students participated in the data collection, of which 49% were male. The schools represented urban (40%), suburban (20%), and rural settings (40%) [23].

The students interacted with the software six times over the 4-week period, supplemented with six teacher-led lessons. Each interaction with UNCHARTED DISCOVERY was 50 minutes long. For this work, only sessions from the first two weeks were considered. During the first two weeks, four quests were available to the students. During the remaining two weeks, the students had access to two quests that were entirely contained within a single set of actions. From a goal recognition standpoint, these quests would necessitate special consideration, since the student is effectively prevented from performing many of the available actions while attempting the quest. The data from the tutorial level were also removed. Observations from all sessions were concatenated to produce one observation sequence for each student.

6 Results

The evaluation of our goal recognition model focused on three principal metrics: F1,

convergence rate, and convergence point. F1 measures the predictive accuracy of the models. The F1 measure is defined as the harmonic mean between precision and recall. However, since our MLN formulae enforce that only one prediction is made for each time step, the multi-class precision and recall are equivalent to the F1 measure. Convergence rate is the percentage of sequences that are eventually classified to the correct goal. Any sequence whose final action is predicted as belonging to the correct goal is said to have converged on the goal. Convergence point measures the percentage of a converged sequence that was observed before the correct goal was consistently predicted. For this measure, a lower number indicates improved performance.

Two models were trained and evaluated on each of the corpora. The baseline model was a previously published model for goal recognition on the OUTBREAK corpus [20]. It was comprised of 13 logical formulae (Fig. 1) with weights learned using Markov: The Beast [22].

The *discovery events* models extended the baseline model in two ways. First, the model was simplified by removing formulae that referenced previous time steps (i.e., formulae 7-12 in Fig. 1); only formulae that included references to the current time step were maintained. These formulae formed the cores of the models for both the OUTBREAK and UNCHARTED DISCOVERY corpora.

In addition, new sets of formulae were added to encode discovery events in each of the new models. The discovery events allowed the models to reference key events students had completed in the narratives. For each discovery event, a new formula was created. Each discovery event was expressed in terms of the predicates described in Section 3. For example, the *Testing* discovery event for the OUTBREAK corpus was defined as $action(t, "Test")$. The *Eating Habits* discovery event is defined as $argument(t, "Eating Habits")$ because it can be triggered by multiple different types of actions. The encoded discovery events were placed within the milestone formulae, which served as indicator nodes to determine if the discovery event had happened in the interaction history.

$$\forall t, g : |\forall t_2 < t : action(t_2, "Worksheet")| \geq 1 \Rightarrow goal(t, g) * w_{13}(g) \quad (1)$$

$$\forall t, g : |\forall t_2 < t : action(t_2, "Test")| \geq 1 \Rightarrow goal(t, g) * w_{14}(g) \quad (2)$$

$$\forall t, g : |\forall t_2 < t : argument(t_2, "Eating Habits")| \geq 1 \Rightarrow goal(t, g) * w_{15}(g) \quad (3)$$

$$\forall t, g : \left| \forall t_2 < t : \begin{array}{l} action(t_2, "Read") \wedge \\ argument(t_2, "Salmonellosis") \end{array} \right| \geq 1 \Rightarrow goal(t, g) * w_{16}(g) \quad (4)$$

$$\forall t, g : |\forall t_2 < t : argument(t_2, "Symptoms")| \geq 1 \Rightarrow goal(t, g) * w_{17}(g) \quad (5)$$

$$\forall t, g : |\forall t_2 < t : argument(t_2, "Bacteria")| \geq 1 \Rightarrow goal(t, g) * w_{18}(g) \quad (6)$$

Fig. 2. Discovery event formulae for CRYSTAL ISLAND: OUTBREAK

$$\begin{aligned} \forall t, g : |\forall t_2 < t : action(t_2, \text{"Pickup Flag"})| \geq 1 &\Rightarrow goal(t, g) * w_{19}(g) & (1) \\ \forall t, g : |\forall t_2 < t : action(t_2, \text{"Drop Flag"})| \geq 1 &\Rightarrow goal(t, g) * w_{20}(g) & (2) \\ \forall t, g : |\forall t_2 < t : action(t_2, \text{"Take Notes"})| \geq 1 &\Rightarrow goal(t, g) * w_{21}(g) & (3) \\ \forall t, g : |\forall t_2 < t : action(t_2, \text{"Take Photo"})| \geq 1 &\Rightarrow goal(t, g) * w_{22}(g) & (4) \\ \forall t, g : |\forall t_2 < t : action(t_2, \text{"Use IslandPedia"})| \geq 1 &\Rightarrow goal(t, g) * w_{23}(g) & (5) \\ \forall t, g : |\forall t_2 < t : action(t_2, \text{"Begin Quest"})| \geq 1 &\Rightarrow goal(t, g) * w_{24}(g) & (6) \\ \forall t, g : |\forall t_2 < t : action(t_2, \text{"Quit Quest"})| \geq 1 &\Rightarrow goal(t, g) * w_{25}(g) & (7) \\ \forall t, g : |\forall t_2 < t : action(t_2, \text{"Check Map"})| \geq 1 &\Rightarrow goal(t, g) * w_{26}(g) & (8) \end{aligned}$$

Fig. 3. Discovery event formulae for CRYSTAL ISLAND: UNCHARTED DISCOVERY

The formulae in Fig. 2 capture the discovery events for CRYSTAL ISLAND: OUTBREAK. The formulae in Fig. 3 capture the discovery events for CRYSTAL ISLAND: UNCHARTED DISCOVERY.

The models were trained using 10-fold cross validation. The folds were formed using the number of students, rather than the number of goal sequences, to ensure independence between folds. The three evaluation metrics (F1, convergence rate, convergence point) were computed across each of the ten evaluations. The four models were compared statistically using one-way analysis of variance. The results of cross validation on the CRYSTAL ISLAND: OUTBREAK corpus can be found in Table 2, and the results for the CRYSTAL ISLAND: UNCHARTED DISCOVERY corpus can be found in Table 3.

7 Discussion

The empirical results suggest that logical encodings of domain-specific discovery events improve performance of goal recognition models. The discovery event models outperformed the previous state-of-the-art baselines on all three metrics for both cor-

Table 2. Goal recognition model comparison for CRYSTAL ISLAND: OUTBREAK

Model	F1		Convergence Rate		Convergence Point	
Baseline	0.488	F(1,18)=7.661 p<0.05	30.906	F(1,18)=133.7 P<0.001	50.865	F(1,18)=27.98 p<0.001
Discovery Event	0.546		50.056		35.862	

Table 3. Goal recognition model comparison for CRYSTAL ISLAND: UNCHARTED DISCOVERY

Model	F1		Convergence Rate		Convergence Point	
Baseline	0.226	F(1,18)=14.58 p<0.01	11.915	F(1,18)=709.1 p<0.001	87.786	F(1,18)=61.66 p<0.001
Discovery Event	0.244		29.973		79.350	

pora. In both cases, the discovery event models were significantly more accurate than the baseline models, as demonstrated by analyses of variance on F1 scores. In addition, the discovery event models consistently generated accurate predictions earlier in the observation sequences. Early prediction is particularly important for goal recognition in narrative-centered learning environments, as it provides time for computing and executing interventions to dynamically shape interactive narratives. In the future, it may be useful to consider alternate measures of early prediction. One example would be to define a sequence as converged if there is high probability that a large proportion of the observations are consistent

It is important to note that there are several limitations to the work. First, there are some restrictions on the types of interactive environments in which this approach can be used. The approach is most compatible with environments that are explicitly goal-oriented. Automatically encoding user progress is a necessary part of defining discovery events. Identifying meaningful progress measures for an environment without well-defined goals is likely to be problematic. It is conceivable that navigating physical landmarks could serve as progress milestones for purely exploration-focused environments, and this type of investigation is left for future work.

A second limitation is the assumption that there are no concurrent goals. Relaxing this assumption would likely require a resource-intensive manual annotation effort to determine how various actions contribute to the goals to be recognized. Alternatively, asking students to periodically self-report their goals—either through the software or think aloud protocol—could also be explored.

Lastly, there is the question of appropriate grain size for discovery events representations. On the one hand, it is possible that a coarse-grained representation that solely denotes which goals have been completed would provide sufficient context for effective goal recognition. On the other hand, every individual action performed by a student provides context for inferring the goal she is trying to accomplish. The discovery events presented in this work were chosen because they provide salient pieces of plot-related information to the student; a retelling of the story would be incoherent without including the information conveyed by the discovery events. It is unclear whether this is necessarily the optimal grain size for encoding discovery events, but ultimately it is an empirical question. In the future, additional studies should be conducted to determine the appropriate representational characteristics for encoding discovery events to drive goal recognition in narrative-centered learning environments.

8 Conclusions and Future Work

Goal recognition has a long history of study in artificial intelligence. In narrative-centered learning environments, goal recognition models enable interactive narrative planners to personalize events to individual students and their problem-solving behaviors. In this paper, we have presented a goal recognition framework that leverages Markov logic networks to encode relations between users' goals and discovery events, key milestones denoting user progress in a narrative-centered learning environment. Discovery event-based models were compared to a previous state-of-the-art

MLN model across two narrative-centered learning environments. Empirical analyses demonstrated that the discovery event-based models consistently outperformed the baseline model on all considered goal recognition metrics: F1 score, convergence rate, and convergence point.

There are several promising directions for future work. First, it will be important to investigate ways to relax the simplifying assumptions used in the work, such as the assumption that whenever a student achieves a goal, it is the same goal that she had been seeking to achieve. Another assumption it will be important to relax is the assumption that goals are pursued serially. In open-ended narrative-centered learning environments, such as both editions of CRYSTAL ISLAND, it is possible for students to pursue multiple goals concurrently, and in some cases synergistically. Finally, it will be important to incorporate discovery event-based goal recognition models into run-time narrative-centered learning environments to investigate their capacity to drive user-adaptive systems.

Acknowledgments. The authors wish to thank members of the IntelliMedia Group for their assistance, as well as Valve Software for access to the Source™ engine and SDK. We would also like to thank Eunyoung Ha, for providing the baseline model. This research was supported by the National Science Foundation under Grant DRL-0822200. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Additional support was provided by the Bill and Melinda Gates Foundation, the William and Flora Hewlett Foundation, and EDUCAUSE.

References

1. Gal, Y., Reddy, S., Shieber, S.M., Rubin, A., Grosz, B.J.: Plan Recognition in Exploratory Domains. *Artif. Intell.* 176, 2270–2290 (2012).
2. Blaylock, N., Allen, J.: Hierarchical Instantiated Goal Recognition. In: Kaminka, G., Pynadath, D. V., and Geib, C.W. (eds.) *Proceedings of the Workshop on Modeling Others from Observations*. pp. 8–15. AAAI Press, Menlo Park, CA (2006).
3. Gold, K.: Training Goal Recognition Online from Low-Level Inputs in an Action-Adventure Game. In: Youngblood, G.M. and Bultko, V. (eds.) *Proceedings of the Sixth Conference on Artificial Intelligence in Interactive Digital Entertainment*. pp. 21–26. AAAI Press, Menlo Park, CA (2010).
4. Kabanza, F., Bellefeuille, P., Bisson, F., Benaskeur, A.R., Irandoust, H.: Opponent Behaviour Recognition for Real-Time Strategy Games. In: Sukthankar, G., Geib, C.W., Pynadath, D. V., and Bui, H.H. (eds.) *Proceedings of the Workshop on Plan, Activity, and Intent Recognition*. pp. 29–36. AAAI Press, Menlo Park, CA (2010).
5. Rowe, J.P., Shores, L.R., Mott, B.W., Lester, J.C.: Integrating Learning, Problem Solving, and Engagement in Narrative-Centered Learning Environments. *Int. J. Artif. Intell. Educ.* 21, 115–133 (2011).
6. Thomas, J.M., Young, R.M.: Annie: Automated Generation of Adaptive Learner Guidance for Fun Serious Games. *IEEE Trans. Learn. Technol.* 3, 329–343 (2010).
7. Domingos, P., Kok, S., Poon, H., Richardson, M., Singla, P.: Unifying Logical and Statistical AI. In: Gil, Y. and Mooney, R.J. (eds.) *Proceedings of the Twenty-First*

- National Conference on Artificial Intelligence. pp. 2–7. AAAI Press, Menlo Park, CA (2006).
8. Kautz, H.A., Allen, J.F.: Generalized Plan Recognition. In: Kehler, T. (ed.) Proceedings of the Fifth National Conference on Artificial Intelligence. pp. 32–37. Morgan Kaufmann, San Francisco, CA (1986).
 9. Ng, H.T., Mooney, R.J.: Abductive Plan Recognition and Diagnosis: A Comprehensive Empirical Evaluation. Third International Conference on Principles of Knowledge Representation and Reasoning. pp. 499–508. , Cambridge, MA (1992).
 10. Geib, C.W., Goldman, R.P.: A Probabilistic Plan Recognition Algorithm Based on Plan Tree Grammars. *Artif. Intell.* 173, 1101–1132 (2009).
 11. Charniak, E., Goldman, R.P.: A Bayesian Model of Plan Recognition, (1993).
 12. Pynadath, D. V., Wellman, M.P.: Probabilistic State-Dependent Grammars for Plan Recognition. In: Boutilier, C. and Goldszmidt, M. (eds.) Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence. pp. 507–514. IJAI Press, Corvallis, OR (2000).
 13. Geib, C.W.: Delaying Commitment in Plan Recognition Using Combinatory Categorical Grammars. In: Boutilier, C. (ed.) Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence. pp. 1702–1707. AAAI Press, Menlo Park, CA (2009).
 14. Wu, T., Lian, C., Hsu, J.Y.: Joint Recognition of Multiple Concurrent Activities using Factorial Conditional Random Fields. In: Pynadath, D. V. and Geib, C.W. (eds.) Proceedings of the Workshop on Plan, Activity, and Intent Recognition. pp. 82–88. AAAI Press, Menlo Park, CA (2007).
 15. Sadilek, A., Kautz, H.: Location-Based Reasoning about Complex Multi-Agent Behavior. *J. Artif. Intell. Res.* 43, 87–133 (2012).
 16. Carberry, S.: Techniques for Plan Recognition. *User Model. User-adapt. Interact.* 11, 31–48 (2001).
 17. Albrecht, D.W., Zukerman, I., Nicholson, A.E., Bud, A.: Towards a Bayesian Model for Keyhole Plan Recognition in Large Domains. In: Jameson, A., Paris, C., and Tasso, C. (eds.) Proceedings of the Sixth International Conference on User Modeling. pp. 365–376. Springer Wien New York, Vienna, New York (1997).
 18. Laviers, K., Sukthankar, G.: A Real-Time Opponent Modeling System for Rush Football. In: Walsh, T. (ed.) Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. pp. 2476–2481. AAAI Press, Menlo Park, CA (2011).
 19. Singla, P., Mooney, R.J.: Abductive Markov Logic for Plan Recognition. In: Burgard, W. and Roth, D. (eds.) Twenty-fifth National Conference on Artificial Intelligence. pp. 1069–1075. AAAI Press, Menlo Park, CA (2011).
 20. Ha, E.Y., Rowe, J.P., Mott, B.W., Lester, J.C.: Goal Recognition with Markov Logic Networks for Player-Adaptive Games. In: Bulitko, V. and Riedl, M.O. (eds.) Proceedings of the Seventh Conference on Artificial Intelligence in Interactive Digital Entertainment. pp. 32–39. AAAI Press, Menlo Park, CA (2011).
 21. Baikadi, A., Rowe, J.P., Mott, B.W., Lester, J.C.: Improving Goal Recognition in Interactive Narratives with Models of Narrative Discovery Events. In: Si, M., Cavazza, M., and Zook, A. (eds.) Proceedings of the Sixth Workshop on Intelligent Narrative Technologies. AAAI Press, Menlo Park, CA (2013).
 22. Riedel, S.: Improving the Accuracy and Efficiency of MAP Inference for Markov Logic. In: McAllester, D.A. and Myllymäki, P. (eds.) Proceedings of the Twenty-Fourth Annual Conference on Uncertainty in Artificial Intelligence. pp. 468–475. IJAI Press, Corvallis, OR (2008).
 23. Lester, J.C., Spires, H.A., Nietfeld, J.L., Minogue, J., Mott, B.W., Lobene, E.V.: Designing Game-Based Learning Environments for Elementary Science Education: A Narrative-Centered Learning Perspective. *Information Sciences.* 264, 4-18 (2014)