Multimodal Goal Recognition in Open-World Digital Games

Wookhee Min[1], Bradford Mott[1], Jonathan Rowe[1], Robert Taylor[1], Eric Wiebe[1], Kristy Elizabeth Boyer[2], James Lester[1]

[1]Center for Educational Informatics, North Carolina State University, Raleigh, NC 27695
[2]Department of Computer & Information Science & Engineering, University of Florida, Gainesville, FL 32601
[1]{wmin, bwmott, jprowe, rgtaylor, wiebe, lester}@ncsu.edu, [2]keboyer@ufl.edu

Abstract

Recent years have seen a growing interest in player modeling to create player-adaptive digital games. As a core player-modeling task, goal recognition aims to recognize players' latent, high-level intentions in a non-invasive fashion to deliver goal-driven, tailored game experiences. This paper reports on an investigation of multimodal data streams that provide rich evidence about players' goals. Two data streams, game event traces and player gaze traces, are utilized to devise goal recognition models from a corpus collected from an open-world serious game for science education. Empirical evaluations of 140 players' trace data suggest that multimodal LSTM-based goal recognition models outperform competitive baselines, including unimodal LSTMs as well as multimodal and unimodal CRFs, with respect to predictive accuracy and early prediction. The results demonstrate that player gaze traces have the potential to significantly enhance goal recognition models' performance.

Introduction

With the objective of creating player-adaptive games that are dynamically customized to individual players, player modeling is becoming the subject of increasing attention. Player modeling aims to identify players' changing cognitive and affective states during gameplay. In addition to player behaviors and gameworld events, player modeling may also be able to leverage multimodal sensor data capturing players' verbal behaviors, non-verbal behaviors, and physiological signals (Yannakakis et al. 2013). A wide range of player-modeling tasks have been investigated, including player affect modeling (Martínez, Bengio, and Yannakakis 2013), plan recognition (Bisson, Larochelle, and Kabanza 2015), intent recognition (Min et al. 2016a), and experience estimation (Burelli, Triantafyllidis, and Patras 2014). Player modeling has broad applications in interactive narrative (Riedl and Bulitko 2013), game design (Yannakakis et al. 2013), dynamic game balancing (Lopes and Bidarra 2011), procedural content generation (Shaker, Togelius, and Nelson 2016), and personalized tutoring in educational games (Mott, Lee, and Lester 2006).

Open-world digital games provide players with self-directed gameplay by allowing free exploration of expansive gameworlds in which players choose their own paths to achieve their goals (Squire 2008). However, the high degree of autonomy granted to players poses significant challenges to game designers who have to design coherent storylines and gameworld events (Min et al. 2014; Riedl and Bulitko 2013). Player modeling addresses this challenge in open-world games by supporting the creation of tailored game content attuned to individual players' cognitive and affective states.

As a primary focus of player-modeling research, goal recognition has been the subject of growing attention (Harrison et al. 2015). Player goal recognition is the task of dynamically identifying the high-level objectives that a player is attempting to achieve based on a variety of evidence that is captured during interactions with a game. Goal recognition in open-world games exhibits significant uncertainty: there are a vast number of possible ways to achieve goals, and players may perform exploratory actions, instead of taking focused goal-directed actions, to familiarize themselves with the gameworld (Ha et al. 2011). This characteristic of open-world digital games yields idiosyncratic action and goal achievement sequences. Thus, it is critical to devise goal recognition models that robustly handle uncertainty.

While player modeling based on multimodal data such as facial action units (Hoegen, Stratou, and Gratch 2017) and skin conductance (Martínez, Bengio, and Yannakakis 2013) has been the subject of growing interest, and some work outside of digital games has investigated multimodal intent recognition, such as using hand motions and gaze data in hand-eye coordination tasks (Razin and Feigh 2017), there has been limited exploration of multimodal

data for goal recognition in digital games. In this paper, we present a multimodal goal recognition framework for open-world digital games. To evaluate the effectiveness of multimodal data for goal recognition, we investigate players' real-time gaze traces along with game event traces, and compare multimodal goal recognition models to those induced utilizing only game event traces. We devise goal recognition models using two machine learning techniques that have demonstrated significant success in many sequence-labeling tasks, long short-term memory recurrent neural networks (LSTMs) (Hochreiter and Schmidhuber 1997) and conditional random fields (CRFs) (Lafferty, McCallum, and Pereira 2001). Our prior work has shown that LSTM-based goal recognition models outperform competitive baseline models, but these models have exclusively utilized gameplay trace data (Min et al. 2016a). In this work, we hypothesize that player gaze traces represent an external manifestation of players' goal-directed cognitive processes and that multimodal LSTM-based goal recognition models outperform baseline approaches.

Related Work

Recognizing users' goals and plans holds great promise for increasing the effectiveness of user-adaptive environments. Plan recognition (Sukthankar et al. 2014) is often formulated as a generalized task of goal recognition because it focuses on inferring plans and goals of observed agents. Plan recognition approaches often require a plan library consisting of all potential agent behaviors to be explicitly provided (Bisson, Larochelle, and Kabanza 2015) or they adopt a planning technique assuming close-to-rational agents (Baker et al. 2009; Ramírez and Geffner 2011). We adopt a corpus-based, statistical goal recognition approach that is well suited to the characteristics of open-world games, in which players take exploratory actions in expansive, virtual gameworlds (Blaylock and Allen 2003; Min et al. 2016a).

Recent work on player modeling has investigated multimodal data sources, including body movement (Burelli, Triantafyllidis, and Patras 2014) and head movement (Shaker et al. 2013) for player experience estimation, facial expressions for opponent modeling (Hoegen, Stratou, and Gratch 2017), and physiological signals such as skin conductance and blood volume pulse for player affect modeling (Martínez, Bengio, and Yannakakis 2013). However, multimodal data sources have not been previously investigated with the objective of enhancing computational models of player plan, activity, and intent recognition.

Deep learning (LeCun, Bengio, and Hinton 2015) has shown particular success in player modeling. Summerville and colleagues (2016) investigated procedural content generation using LSTMs that learn latent play styles from gameplay videos, generating levels and predicted paths. Harrison and colleagues (2017) investigated encoder-decoder LSTMs for *AI rationalization*, which translates agents' state-action representations into natural language describing the rationales behind agents' behaviors. Martínez and colleagues (2013) examined convolutional neural networks to model player affect using physiological signals collected during gameplay with a 3D prey-predator game.

Figure 1: The CRYSTAL ISLAND open-world educational game.

In our previous work, we introduced an LSTM-based goal recognition architecture featuring distributed action representations, which significantly outperforms previous state-of-the-art approaches based on deep feedforward neural networks (Min et al. 2014) and Markov logic networks (Ha et al. 2011) with respect to predictive accuracy (Min et al. 2016a). In this paper, we explore an LSTM-based approach to goal recognition that leverages multimodal game trace data. We also evaluate the performance of multimodal goal recognition with more granular goals than was used in prior work.

Multimodal Goal Recognition Framework

In our multimodal goal recognition framework, the inputs to our models consist of low-level game event trace data and player eye gaze trace data. The outputs are the set of possible goals that the player may next achieve in the gameworld. We describe the key components of the goal recognition framework, which we test using the CRYSTAL ISLAND game environment below.

CRYSTAL ISLAND Goal Recognition Testbed

CRYSTAL ISLAND (Figure 1) is an open-world educational game implemented with the Unity game engine for middle school science and literacy education. CRYSTAL ISLAND's gameplay is similar to many exploration-centric games in which players experience the world from a first-person viewpoint and perform actions such as navigating from one location to another, discovering important items, and talking with non-player characters (NPCs). In the game, players are tasked with identifying the cause of an illness

afflicting a team of scientists on a remote island research station. Due to the nature of open-world games, players' behavior sequences do not necessarily represent optimal paths for achieving goals. Players typically make gradual progress toward each objective, eventually culminating in the final actions that solve the science mystery (Min et al. 2016a).

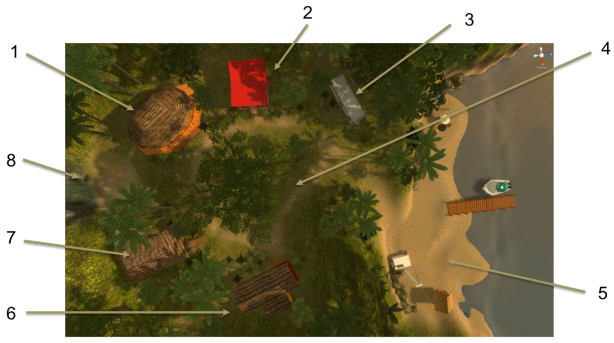
During gameplay, CRYSTAL ISLAND logs all player actions, which can be retrieved for offline data analysis. The data used in the evaluation of the multimodal goal recognition models was collected during a series of data collections for a study involving 140 players from two public middle schools in the United States. Before starting the game, students were given a brief overview of CRYSTAL ISLAND. They played the game during science class over three consecutive days. Prior to each game play session they undertook a brief calibration activity with the eye tracker affixed to their laptop. In the dataset, trace data from 118 players contain eye gaze logs as well as game event logs, while 22 players' data only contain game event logs because there were an insufficient number of eye trackers to provide one for each laptop in the study. In this work, we use data from all 140 players regardless of whether it contains gaze logs because goal recognition models do not need to limit themselves to specific types of data.

Multimodal Inputs

In this paper, we focus on two data channels: game event traces and player gaze traces.

Game event traces. Game event traces record sequences of actions that are triggered by either the player or an NPC, where the former consists of players' in-game behaviors and the latter consists of gameworld events that the player encounters. In our previous work (Min et al. 2014), a player action was encoded with five action properties: action type, action arguments, location, narrative state (an indication of whether a player achieved a set of pre-defined key milestone events within the narrative scenario), and previously achieved goals. In this work, we exclude narrative state because it requires domain-specific knowledge about a game's narrative, which would have hindered generalizability of the goal recognition framework. The remaining four properties are the following:

- **Action Type**: The type of current action taken by the player, such as "*Move*" to a particular location, and "*Talk*" to an NPC. Our data includes 9 distinct types of player actions.
- **Action Arguments**: Arguments that an action type is associated with, such as (EX_1) "*Player*" moves to "*Infirmary*" and (EX_2) "*Camp Nurse*" talks to "*Player*" about "*Spreading Illness*", where the number of action arguments depends on the action type. In this work, we consider two arguments, the *actor* (e.g., Player in EX_1 and Camp Nurse in EX_2) and *target* (e.g., Infirmary in EX_1 and Player in EX_2). Our data, in sum, includes 14 and 180 distinct values for the actor and target arguments, respectively.
- **Location**: The location in the gameworld, where a current player action is taken. It can be either a grid-based coordinate or discretized region within the virtual world. Our data includes 24 non-overlapping, discrete regions that decompose the eight major camp locations (Figure 2).
- **Previously Achieved Goals**: An encoding of the previous goals achieved by the player. A vector initialized with "None" values is populated. Vector elements are updated sequentially as the player achieves goals.



1. Living Quarters
2. Infirmary
3. Laboratory
4. Large Outdoor Region
5. Beach
6. Dining Hall
7. Lead Scientist's Quarters
8. Waterfall

Figure 2: Map of the CRYSTAL ISLAND research camp.

Player gaze traces. As eye trackers have become more affordable and integrated into laptops and monitors, a growing number of digital games (e.g., *For Honor* and *Assassin's Creed® Origins*) are adopting players' eye movements as game input. In the evaluation study, we used Tobii EyeX eye tracking sensors (Figure 3, Top), which use near-infrared light to track the eye movements and gaze points of a player.

To collect eye gaze data during gameplay, we implemented a novel gaze-target-labeling module to enable CRYSTAL ISLAND to automatically process eye tracking data at runtime and identify which objects the player is looking at in the 3D virtual world. The module utilizes ray casting to automatically detect virtual objects that the player fixates upon in the 3D environment. For example, the bottom of Figure 3 shows a fixation on an NPC. Fixations are timestamped and recorded in the trace data with the "Look" action type, an actor argument of "Player," and a target argument of the gaze object. "Look" actions are logged when a fixation event lasts longer than 250 milliseconds, a threshold that is based upon previous research on eye fixations during reading (Rayner 1998). In this manner, the gaze-target-labeling module produces a compact data stream of gaze event data, and it mitigates the

Figure 3: (Top) Students' gameplay. The yellow dashed box is the Tobii EyeX eye tracker used to capture player gaze traces. (Bottom) Gaze label indicated with green eye icon. The game logs the gaze target's name in the game trace data. This screenshot was captured for demonstration purposes. Gaze labels are not displayed during gameplay.

need for extensive feature engineering with coordinate-based eye gaze data.

Since both game event traces and player gaze traces are represented in action properties, we use *actions* to denote both game events and player gaze traces in the following sections of this paper.

Goal Recognition Corpus Generation

The first step in training multimodal goal recognition models is to create a goal recognition corpus. In this subsection, we describe input and output encodings for LSTM and CRF-based goal recognition models.

A key capability provided by deep learning-based machine-learning techniques is automatically extracting hierarchical, multi-level features from data (LeCun, Bengio, and Hinton 2015). Along with multiple levels of nonlinear transformations, deep learning techniques can represent data in a continuous, distributed vector space via a linear projection layer (Figure 4A), which is an approach that has been widely investigated in natural language processing (Mikolov, Yih, and Zweig 2013). Compared to vectors induced by one-hot encoding, which are inherently sparse and high-dimensional (i.e., the size of the vector equals the number of possible values for the variable), distributed

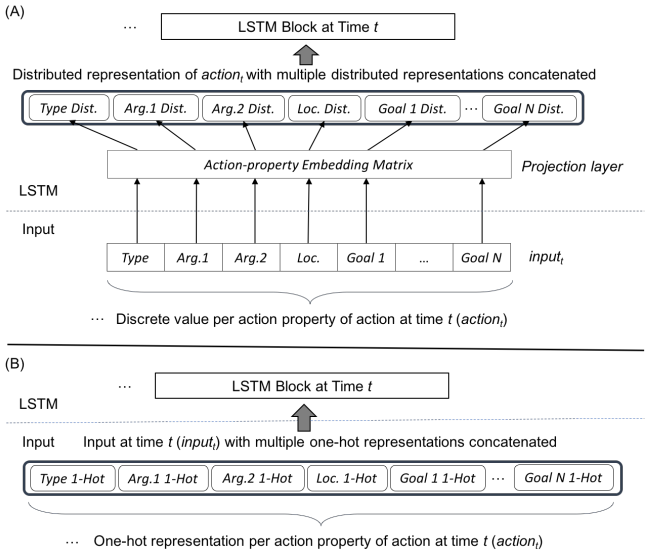


Figure 4: (A) LSTM incorporating a projection layer to learn distributed representations of actions (Min et al. 2016a). The number of input features for a single action is N+4, (B) LSTM with one-hot encoded action inputs. The number of input features is 227+13*N, where 13 is derived from 12 possible goals plus "None", and 227 are the total number of possible values for the other action properties. In both figures, N is the number of previously achieved goals to be considered by the models. In this work, it is set to 12.

representation-based encodings have a representational benefit by efficiently encoding inputs, providing computational efficiencies (Turian, Ratinov, and Bengio, 2010).

We investigate two input representation techniques for LSTM-based goal recognition: (1) distributed representations learned through linear projection supported by an Action-property Embedding Matrix (Figure 4A) (Min et al. 2016a), and (2) one-hot representations (Figure 4B). Since CRF does not support representation learning by itself, we use a one-hot encoding for the input representation in CRF-based goal recognition models.

In the version of CRYSTAL ISLAND utilized in the classroom study, we define 12 goals (i.e., output labels) following (Rowe et al. 2011), where 11 goals are associated with the game's nonlinear narrative and 1 goal is associated with players' off-task behaviors, which are particularly important in educational games (Sabourin et al. 2013). Seven of the narrative-based goals involve speaking with NPCs about the spreading illness, patients' symptoms, food items that may be transmitting the disease, and microbiology concepts. Two narrative-based goals involve (1) testing food items in the virtual laboratory, and (2) discovering the food-based transmission source of the disease through a positive lab test. The two-remaining narrative-based goals involve (1) submitting a diagnosis to the camp nurse for the first time, and (2) submitting a correct diagnosis that solves the mystery.

Table 1 presents descriptive statistics on the goal recognition corpus based on 140 players' gameplay traces. Note that we use "Look" actions as predictors for multimodal goal recognition, but goal predictions are made only for non-Look actions, since unimodal goal recognition models do not consider Look actions. In this manner, the number of goal predictions between multimodal and unimodal goal recognition models are the same. Table 1 reports statistics excluding Look actions. The highest number of player actions associated with accomplishing a goal are for "Test Contaminated Object (Goal A)," which serves as a majority-class baseline of 19.1%. The lowest number of actions was associated with the "Talk to Camp Nurse (Goal B)," which is typically the first goal that the player accomplishes in the game.

Total Number of Observed Actions	93,844
Total Number of Goals Achieved	1,287
Total Number of Actions Labeled as Goal A	17,916
Total Number of Actions Labeled as Goal B	1,872

Table 1: Descriptive statistics of the goal recognition corpus.

Goal Recognition Model Training

We cast goal recognition as a multiclass classification problem in which a trained classifier predicts the most likely goal associated with the currently observed action sequence (Min et al. 2014). In this work, we investigate two sequence-labeling techniques: LSTMs and CRFs. LSTMs are a variant of recurrent neural networks that feature three gating units to adaptively maintain long-term memory in time-series data and alleviate vanishing and exploding gradient problems in model training (Hochreiter and Schmidhuber 1997). CRFs are discriminative, undirected graphical models, which are specifically designed to learn interdependencies among output variables for structured prediction (Lafferty, McCallum, and Pereira 2001). In this work, we examine single-layer LSTMs and linear-chain CRFs.

We evaluate LSTM-based goal recognition using unimodal (i.e., game trace) and multimodal (i.e., game trace + eye gaze trace) data and different input encoding methods (distributed action representations vs. one-hot encoding) along with CRF-based competitive baselines. Both for LSTMs and CRFs, a set of hyperparameters are chosen prior to training models. We adopt a grid search method using 10-fold cross-validation (100 players' data), and the best performing model configurations are evaluated on a held-out test set (40 players' data). Hyperparameters explored for LSTMs include the two input encoding methods—one-hot encoding (one.) and distributed representations (dist.)—and the maximum length of sequences in the input among {50, 100}. CRFs also explore the maximum length of sequences in the input among {50, 100}, along with the maximum number of iterations to find constraints and perform updates for models among {10, 20, 30, 40}.

For the LSTM models, we use an action-property embedding size of 20 (only for the distributed representation setting), 100 hidden units, and a dropout rate of 0.75 (Srivastava et al. 2014). We adopt a mini-batch gradient descent with a mini-batch size of 128, and we utilize categorical cross entropy for the loss function and the Adam stochastic optimizer (Kingma and Ba 2015). Finally, the training process stops early if the validation score has not improved within the last seven epochs. In each fold, 10% of the training data is used to determine early stopping, and 90% is utilized for supervised training, while the validation data is purely used for calculating the validation score. The maximum number of epochs is set to 100. For CRFs, we use a structured support vector machine solver using the 1-slack formulation and cutting plane method (Joachims, Finley, and Yu 2009) for model optimization and a regularization parameter of 1.

Evaluation

To evaluate the predictive capabilities of the multimodal goal recognition models, two input feature sets are designed: (1) unimodal, which uses game event traces only (U), and (2) multimodal, which uses both game event traces and gaze traces (M). As noted, since goal predictions are made on all non-Look actions, models trained on U and M have the same number of data instances for both the training set (100 players containing 87 players with gaze data) and held-out test set (40 players containing 31 players with gaze data), respectively.

We conducted 10-fold cross-validation for model hyperparameter optimization based on the training set. In cross-validation, we used the same player-level data split and performed a grid search on hyperparameters, and chose the best performing model configurations based on accuracy rate only. During testing, models' early prediction capacity was measured with *standardized* convergence point (SCP) and *n*-early convergence rate metrics (*n*-CR) (Min et al. 2016b) for hyperparameter-optimized models based on the held-out test set. We also calculated models' accuracy rates during testing.

SCP measures how early goal recognition results converge to a correct goal within an action sequence while penalizing non-converged action sequences (e.g., the last action's goal prediction in an action sequence is incorrect). In contrast to other metrics for early prediction, such as convergence point (Blaylock and Allen 2003), SCP measures early prediction capacity by calculating scores less than 1 for converged sequences and scores greater than 1 for non-converged sequences. The penalty parameter in SCP should be determined considering the game's charac-

teristics. Based on our corpus, the average number of non-Look actions to achieve a goal is 73, which constitutes a sizable amount of gameplay time. To penalize goal recognition models' long-term inefficiency on non-converged sequences, we set the parameter value to "4 times the total number of actions in the sequence," which results in every non-converged sequence's SCP of 5.

n-CR measures if goal predictions on the last *n*+1 actions in an action sequence are consistently correct. Due to space limitations, we only report *n* of 0 and 1 for *n*-CR, by which we evaluate whether goal recognizers correctly predict the goal for the last one and two actions in every action sequence. Lower is better for SCP, and higher is better for *n*-CR.

CRF	U	M	LSTM	U	M
{10-50}	26.30	27.88	{Dist.-50}	36.97	35.55
{20-50}	27.71	30.64	{One.-50}	37.55	**38.49**
{30-50}	28.84	31.77	{Dist.-100}	34.68	33.72
{40-50}	30.78	29.29	{One.-100}	**38.14**	34.05
{10-100}	22.33	22.98			
{20-100}	27.69	24.89			
{30-100}	28.19	**32.86**			
{40-100}	**32.08**	29.69			

Table 2: Averaged 10-fold cross-validation results on predictive accuracy (%). The convention used in CRFs and LSTMs is {*number of iterations-maximum sequence length*} and {*encoding-maximum sequence length*}, respectively. The best accuracy rate per algorithm paired with a feature set is marked bold.

Table 2 presents predictive accuracy rates of the best performing LSTMs and CRFs based on the two feature sets in cross-validation. Overall, the highest accuracy rate (38.49%) is achieved by LSTMs with a one-hot encoding and a maximum sequence length of 50 when trained using both the game event traces and player gaze traces. It is noteworthy that both LSTMs and CRFs (32.86%) achieve higher cross-validation accuracy rates when utilizing both modalities, and they also significantly outperform the majority class-based baseline (19.1%).

Based on the optimized hyperparameter sets for LSTM-U/M and CRF-U/M, we further evaluate models' generalization performance on the held-out test set. In this step, all four models are retrained using all training data (i.e., 100 players' data). Table 3 reports test set-based evaluation results.

Test results indicate that the multimodal LSTM-based goal recognition model attains the highest predictive accuracy with a marginal improvement of 6.7% over the second-best model (unimodal LSTM), which shows LSTMs are also generalizable to unseen player data. It is notable that this multimodal approach outperforms the unimodal LSTM model as well as the CRF-based competitive baseline models with respect to all the targeted metrics. It should be noted that the multimodal CRF-based goal recognition model exhibits sizable decreases in accuracy rate (32.86% to 27.17%) in the generalization test. These findings suggest that the multimodal, linear-chain CRF is not an efficient goal-modeling approach in this domain. The accuracy rates reported in (Min et al. 2016a) are higher relative to this work; however, the previous unimodal work was based on a different version of CRYSTAL ISLAND that used a different goal recognition corpus and a smaller number of goals (seven goals) compared to the 12 fine-grained goals in this work.

	CRF-U	CRF-M	LSTM-U	LSTM-M
Accuracy Rate	33.69	27.17	37.15	**39.65**
SCP	304.73	333.48	302.33	**298.56**
0-CR	41.81	35.59	42.09	**44.07**
1-CR	39.27	32.77	40.96	**42.09**

Table 3: Test set evaluation results (%) for hyperparameter-optimized goal recognition models. The best score per metric is marked bold.

Conclusion

Goal recognition is a key component of player modeling. Multimodal data streams show significant promise as sources of evidence for goal recognition models to deal with the significant uncertainty inherent in player modeling in open-world digital games. We have presented a multimodal goal recognition framework that analyzes bimodal data channels—player gaze traces along with game event traces—for open-world digital games. Empirical results indicate that multimodal LSTM-based goal recognition models achieve high performance on both predictive accuracy and early prediction. Player gaze traces serve as an external manifestation of goal-directed cognitive processes, complementing game event traces that have been a traditional source of input for player goal recognition. In the future, it will be important to investigate additional modalities, including facial expressions, body movements, posture, and biometrics. In addition, it will be important to investigate how multimodal goal recognition models support game adaptations at runtime.

Acknowledgements

This research was supported by the National Science Foundation under Grant CHS-1409639. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Baker, C., Saxe, R., and Tenenbaum, J. 2009. Action understanding as inverse planning. *Cognition* 113(3):329–349.

Bisson, F., Larochelle, H., and Kabanza, F. 2015. Using a Recursive Neural Network to Learn an Agent's Decision Model for Plan Recognition. In *Proceedings of IJCAI*, 918–924.

Blaylock, N., and Allen, J. 2003. Corpus-based, statistical goal recognition. In *Proceedings of IJCAI*, 1303–1308.

Burelli, P., Triantafyllidis, G., and Patras, I. 2014. Non-invasive player experience estimation from body motion and game context. In *Proceedings of IEEE Conference on Computational Intelligence and Games*, 92–98.

Ha, E. Y., Rowe, J. P., Mott, B. W., and Lester, J. C. 2011. Goal Recognition with Markov Logic Networks for Player-Adaptive Games. In *Proceedings of AIIDE*, 32–39.

Harrison, B., Ware, S., Fendt, M., and Roberts, D. 2015. A Survey and Analysis of Techniques for Player Behavior Prediction in Massively Multiplayer Online Role-Playing Games. *IEEE Transactions on Emerging Topics in Computing* 3(2):260–274.

Harrison, B., Ehsan, U., and Riedl, M. O. 2017. Rationalization: A Neural Machine Translation Approach to Generating Natural Language Explanations. *arXiv preprint arXiv:1702.07826*.

Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural Computation* 9(8):1–32.

Hoegen, R., Stratou, G., and Gratch, J. 2017. Incorporating Emotion Perception into Opponent Modeling for Social Dilemmas. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, 801–809.

Joachims, T., Finley, T., and Yu, C. J. 2009. Cutting-plane training of structural SVMs. *Machine Learning* 77(1):27–59.

Kingma, D. P., and Ba, J. L. 2015. Adam: a Method for Stochastic Optimization. In *Proceedings of the International Conference on Learning Representations*.

Lafferty, J., McCallum, A., and Pereira, F. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the International Conference on Machine Learning,* 282–289.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. Deep Learning. *Nature* 521(7553):436–444.

Lopes, R., and Bidarra, R. 2011. Adaptivity challenges in games and simulations: A survey. *IEEE Transactions on Computational Intelligence and AI in Games* 3(2):85–99.

Martinez, H. P., Bengio, Y., and Yannakakis, G. 2013. Learning Deep Physiological Models of Affect. *IEEE Computational Intelligence Magazine*, 8:20–33.

Mikolov, T., Yih, W., and Zweig, G. 2013. Linguistic regularities in continuous space word representations. In *Proceedings of NAACL-HLT*, 746–751.

Min, W., Ha, E. Y., Rowe, J., Mott, B., and Lester, J. 2014. Deep Learning-Based Goal Recognition in Open-Ended Digital Games. In *Proceedings of AIIDE*, 37–43.

Min, W., Mott, B., Rowe, J., Liu, B., and Lester, J. 2016a. Player Goal Recognition in Open-World Digital Games with Long Short-Term Memory Networks. In *Proceedings of IJCAI*, 2590–2596.

Min, W., Baikadi, A., Mott, B., Rowe, J., Liu, B., Ha, E. Y., and Lester, J. 2016b. A Generalized Multidimensional Evaluation Framework for Player Goal Recognition. In *Proceedings of AIIDE*, 197–203.

Mott, B., Lee, S., and Lester, J. 2006. Probabilistic goal recognition in interactive narrative environments. In *Proceedings of the National Conference on Artificial Intelligence*, 187–192.

Ramírez, M., and Geffner, H. 2011. Goal recognition over POMDPs: Inferring the intention of a POMDP agent. In *Proceedings of IJCAI*, 2009–2014.

Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124(3):372–422.

Razin, Y., and Feigh, K. 2017. Learning to Predict Intent from Gaze During Robotic Hand-Eye Coordination. In *Proceedings of AAAI*, 4596–4602.

Riedl, M. O., and Bulitko, V. 2013. Interactive Narrative: An Intelligent Systems Approach. *AI Magazine* 34(1):67–77.

Rowe, J., Shores, L., Mott, B., and Lester, J. 2011. Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education* 21(1–2):115–133.

Sabourin, J. L., Rowe, J. P., Mott, B. W., and Lester, J. C. 2013. Considering alternate futures to classify off-task behavior as emotion self-regulation: A supervised learning approach. *Journal of Educational Data Mining* 5(1):9–38.

Shaker, N., Togelius, J., and Nelson, M. 2016. *Procedural Content Generation in Games*, Springer.

Shaker, N., Asteriadis, S., Yannakakis, G., and Karpouzis, K. 2013. Fusing visual and behavioral cues for modeling user experience in games. *IEEE Transactions on Cybernetics* 43(6):1519–1531.

Squire, K. (2008). Open-Ended Video Games: A Model for Developing Learning for the Interactive Age. *The Ecology of Games: Connecting Youth, Games, and Learning* 167–198.

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. 2014. Dropout : A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research* 15:1929–1958.

Sukthankar, G., Geib, C., Bui, H., Pynadath, D., and Goldman, R. P. 2014. *Plan, Activity, and Intent Recognition*. Elsevier.

Summerville, A., Guzdial, M., Mateas, M., and Riedl, M. O. 2016. Learning player tailored content from observation: Platformer level generation from video traces using LSTMs. In *Proceedings of AIIDE Workshop on Experimental AI in Games*, 107–113.

Turian, J., Ratinov, L., and Bengio, Y. 2010. Word Representations: A Simple and General Method for Semi-supervised Learning. In *Proceedings of the Annual Meeting of the ACL*, 384–394.

Yannakakis, G. N., Spronck, P., Loiacono, D., and André, E. 2013. Player modeling. In *Dagstuhl Follow-Ups (Vol. 6)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.