Detecting Player Goals from Game Log Files

Kristen E. DiCerbo
Pearson
kristen.dicerbo@pearson.com

Khusro Kidwai
University of Southern Maine
khusro.kidwai@maine.edu

ABSTRACT

In this paper we describe the development of a detector of *seriousness of pursuit* of a particular goal in a digital game. As gaming researchers attempt to make inferences about player characteristics from their actions in open-ended gaming environments, understanding game players' goals can help provide an interpretive lens for those actions. This research uses Classification and Regression Tree methodology to develop and then cross-validate features of game play and related rules through which player behavior about pursuing a goal of completing a quest can be classified as serious or not serious.

Keywords

Detector, goal, stealth assessment, games, educational data mining.

1. INTRODUCTION

Many recently-developed online learning environments provide open spaces for students to explore. At the same time, there is growing interest in stealth assessment [5], or the use of data resulting from students' every day interactions in a digital environment to make inferences about player characteristics.

This use of data from natural activity in open-ended environments presents a challenge for interpretation. Much of the evidence we wish to use to assess skill proficiency and player attributes assumes that individuals are working towards the goal of completion of sub-tasks or levels within a game. However, game players often appear to be pursuing differing goals [2], which provide different lenses for interpreting player behavior based on data in log files for games. For example, behavior might be categorized as "off-task" if a player is pursuing a quest but "on-task" if a player has a goal of exploring the environment. If we are interested in using evidence contained in game log files to assess constructs such as persistence, we have to be careful not to identify a player as lacking persistence when in fact they were very persistently pursuing a different goal.

This paper describes the creation of a detector for a specific goal—serious pursuit of completion of quests in a game. The approach taken in this paper builds on research regarding detectors for gaming the system [1], which use machine learning to identify features and rules to classify behavior into discrete categories. In this paper, the focus was on whether a player in the online game Poptropica® is seriously pursuing the goal of completing quests in the game. The ability to successfully categorize players based on whether or not they are pursuing the goal of quest completion is likely to help with interpretation of other actions players pursue in the game. This paper discusses the selection of possible features or indicators of goal seriousness, the process of detector creation, and the analysis of the effectiveness of the detector in correctly classifying play.

2. DESCRIPTION OF THE ENVIRONMENT

Poptropica® is a virtual world in which players explore "islands" with various themes and overarching quests that players can choose to pursue. Players choose which islands to visit and the quests generally involve completion of 25 or more steps (for example, collecting and using assets) that are usually completed in a particular order. Apart from the quests, players can talk to other players in highly scripted chats (players can only select from a pre-determined set of statements to make in the chat sessions), play arcade-style games head-to-head, and spend time creating and modifying their avatar.

Like with most online gaming environments, the Poptropica® gaming engine captures time-stamped event data for each player. On an average day actions of over 350,000 Poptropica® players generate 80 million event lines.

3. DETECTOR DEVELOPMENT

Prior to building a machine detector of goal-seriousness, it was necessary to establish a human-coded standard from which the computer could learn and verify rules. A total of 527 clips were coded by two raters as being either "serious" or "not serious" about the goal of completing a quest. Cohen's Kappa [3] between the two raters for the full set of non-training clips was .72; all disagreements were discussed until accord was reached.

Elements of the log files hypothesized to be indicative of goal directedness were identified as features including: (1) total number of events completed on the island, (2) total amount of time spent on the island, (3) total number of events related to quest-completion, (4) number of locations (scenes) visited on the island, (5) number of costumes tried on, and (6) number of inventory checks. The number of costumes and number of inventory checks were hypothesized to be negatively correlated to completing quests.

Researchers employed a Classification and Regression Tree (CART) methodology to create the detector. The process of the creation of decision trees begins with the attempt to create classification rules until the data has been categorized as close to perfectly as possible, however, this can result in overfit to the training data. The software then tries to "relax" these rules, in a process called "pruning" to balance accuracy and flexibility to new data. This research employed the J48 algorithm [4] for pruning. The results of the analyses were evaluated using (1) precision, (2) recall, (3) Cohen's kappa, and (4) A'.

4. RESULTS

The final decision tree is displayed in Figure 1. Each branch provides classification rules and an ultimate classification decision at the end. The red boxes end paths that indicate individuals non-serious about the goal of quest completion while the green boxes indicate serious goal directed behavior. So, for example, following the left-most path, we find that people who visit 4 or fewer scenes and complete 2 or fewer quest events were classified as not seriously goal-directed. This rule correctly classified 335 clips and misclassified 18 of the 353 total clips that followed this pattern. Following other branches reveals different rules, all leading to classifications of seriously or not seriously goal-directed.

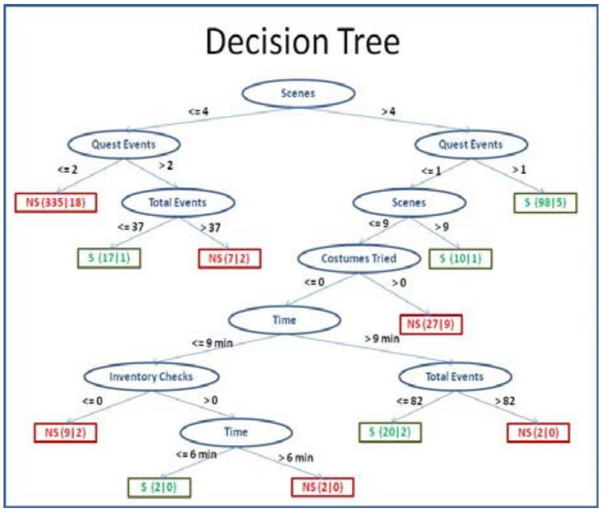

Figure 1. Final Decision Tree

Cross-validation was completed by having half the sample serve as the training sample and the other half as the test sample, and then switching the halves. The detector achieved good performance under cross-validation. Human raters identified 167 of the 527 clips as indicating serious goal-directed behavior. The detector identified 151 clips as serious, out of which 119 agreed with the human raters and 32 did not (see Table 1). This resulted in a precision score of .79 and a recall score of .71. The Kappa value was .63, indicating that the accuracy of the detector was 63% better than chance. The A' was .93, indicating that the detector could correctly classify whether a clip contained serious goal-directed behavior 93% of the time.

Table 1. Correctness of Detector Classification

		Detector	
		Not Serious	Serious
Human	Not Serious	328	32
	Serious	48	119

In order to further investigate where the detector had difficulty with accuracy, we compared places where the detector disagreed with the raters to places where the raters had initially disagreed with each other. Although the human raters eventually came to an agreement about the classification of the clip, their initial disagreement was likely an indication of an ambiguous clip. As displayed in Table 2, when the human raters agreed, the detector also agreed with them 89% of the time. However, when the human raters disagreed, the detector disagreed with their final rating 49% of the time.

Table 2. Cross-tabulation of computer and human agreement

	Humans Disagree	**Humans Agree**
Computer/ Humans Disagree	49.15% (N=29)	10.90% (N=51)
Computer/ Humans Agree	50.85% (N=30)	89.10% (N=417)

5. DISCUSSION AND CONCLUSIONS

The purpose of this paper was to describe the creation of a detector of seriousness of goal pursuit in an online game. The goal a player is pursuing in a game or other open-ended online environment can help provide context and important interpretation of player actions within the system. The results here suggest that an automated detector can be created that can reliably identify whether a participant is seriously pursuing a goal of completing a game quest. The methodology discussed in this paper opens up the possibility of gaming engines detecting and prompting players to adjust their approach (for example, becoming more goal directed) in real time.

We suggest that this type of analysis of player goals may have advantages over other attempts to measure and assess constructs based on player behavior captured in log files. Algorithms such as the one discussed in this paper allow for efficient categorization of thousands of players and millions of actions that would not be humanly feasible. In other games, a similar process could be followed by 1) identifying potential goals, 2) identifying potential indicators of those goals, 3) hand coding a small set of log files as pursuing on of those goals or not, and 4) carrying out the Classification and Regression Tree analysis. The identification of player goals can help us understand player actions in games and extend our ability to make inferences about player characteristics.

6. ACKNOWLEDGMENTS

Our thanks to Ryan Baker and Ilya Golden who served as mentors during this project's inception at the Pittsburgh Science of Learning Center LearnLab Summer School 2012.

7. REFERENCES

[1] Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., and Roll, I. 2006. Generalizing detection of gaming the system across a tutoring curriculum. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 402-411.

[2] Bartle, R. 1996. Hearts, clubs, diamonds, spades: Players who suit MUDs. J.MUD Res, 1, 19.

[3] Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 20(1), 37-46.

[4] Quinlan, R. 1993. C4.5 Programs for Machine Learning. Morgan Kaufman, San Mateo, CA.

[5] Shute, V. 2011. Stealth assessment in computer-based cames to support learning. In *Computer Games and Instruction*, S. Tobias & D. Fletcher, Eds. Information Age Publishers, Charlotte, NC, 503-524.