Evaluating a Drama Management Approach in an Interactive Fiction Game

Andrea Corradini

Institute of Business Communication and Information
Science
University of Southern Denmark
Kolding, Denmark
andrea@sitkom.sdu.dk

Manish Mehta, Santi Ontañón

Cognitive Computing Lab
Georgia Institute of Technology
Atlanta, Georgia, USA
{mehtama1,santi}@cc.gatech.edu

Abstract — **In this paper, we evaluate a drama management approach deployed in an implementation of a graphical interactive fiction game. Our approach uses players' feedback as a basis for guiding the personalization of the interaction. Twenty subjects, with different levels of expertise, took part in the experimental evaluation. The results indicate that our Drama Manager (DM) helps in providing a better play experience. Quantitative analysis indicates that the DM causes an average subjective improvement of 12.5% of the subjects' play experience. Qualitative analysis from the user interview shows that all subjects noticed the difference in playing the game with and without the DM while the former way of playing was also the preferred one. The large majority of players found the hints, and ultimately the DM, helpful even if frustrating at times when these cannot be exploited directly for further proceeding in the game.**

Keywords-component; interactive fiction game; evaluation; drama management

I. INTRODUCTION AND RELATED WORK

There has been a growing interest in creating story-based interactive fiction games where the player is considered an active participant in the ongoing narratives. The component in charge of guiding the complete dramatic experience is called Drama Manager (DM) [2] or Director [5]. The DM employs a set of actions provided at appropriate points in the ongoing game whereby the player is guided toward certain aspects of the story.

In a previous work [7], we evaluated search-based DM techniques in a simple text based implementation of the game Anchorhead [3]. We have developed our own graphical version that incorporates the results from our previous work with the addition of several architectural enhancements [10]. We have tackled the problem of using search-based DM techniques in real-time games and have applied our DM techniques to a game where the number of actions the player can execute is very large. In this paper, we focus on the evaluation of our new DM approach situated in the Anchorhead game.

The idea of treating drama management as an optimization problem was proposed in [2, 8]. However the DMs employed were not connected to a concrete game and the techniques were tested using simulated players. Players' preferences were also not taken into account. To optimize DM actions in response to player actions, reinforcement learning was exploited in [6]. The central premise of this approach is to provide the author with the ability to specify what constitutes a good story. A simulated player model to predict the next player action is used. Yet, no player preference model to measure the interestingness of a story from the player's perspective is built. The Mimesis architecture [9] proposes a story planning based approach for real-time virtual worlds. Story plans are tagged with causal structure and the system handles player actions that might threaten the causal links through either re-planning the story or disallowing the player the opportunity to carry out his action. Only the author can specify the goals that the planner should achieve but no player preference model is incorporated.

In our approach to drama management, we build a player preference model based on real human player interactions with the game. Previous approaches employed only an author based evaluation function for story interestingness.

II. USABILITY STUDY: METHODOLOGY

We carried out the study with twenty participants: eleven males and nine females. Henceforth, we will refer to a certain subject by means of a label P1 through P20 where PN indicates subject N. The subjects were recruited among the employees and graduate students of a local University in Kolding, Denmark. The participants' average age was 31.5 years. Subjects were of several nationalities, all with advanced skills in English. Forty percent of them judge themselves as experienced videogame/console players.

Each user evaluation consisted of 2 different sessions. The first session had an average duration of 60 minutes. At first, we collected and analyzed data about the user and his/her play style, preferences, previous gaming experience and favorite game genres. While user analysis has typically been restricted to demographics and general preferences, we took into account also the users' general knowledge of the task domain and of the system to stress the importance of users' knowledge to their interactions with the game. Immediately thereafter, every player was provided with an explanation of the Anchorhead game. Following a five minute warm-up session the actual playing session lasting up to 25 minutes was started. Eventually, a post-test interview terminated the user testing. A second interactive session lasted an average of 40 minutes and consisted of another playing time of up to 25 minutes and with different settings

to those employed in the first gaming session. At the end of the game time, one more post session interview was briefly carried out.

For the actual playing sessions, we uniformly split by gender the set of participants into 4 groups of 5 people. Henceforth, we refer to each of these groups as G1, G2, G3, and G4, respectively. Each group of people played the game under different conditions regarding the presence/absence of the DM as well as the use of previously collected games. Subjects were not told anything about the game settings. Procedurally, subjects played according to the following guidelines:

- Initially, subjects P1 to P5, belonging to group G1, played the game without the DM. We collected both the game log files and players' feedback at the end of each game. We refer to the collected log files as CG1 game cases.

- Then, subjects P6 to P10, belonging to group G2, were first invited to play with the DM activated and utilizing the game cases CG1. Later they played without the DM. We collected both the game log files and players' feedback at the end of each game session. We refer to the collected log files as the game cases CG2.

- Subjects P11 to P15, belonging to group G3, first played without the DM and then with it and utilizing the game cases CG1. We collected both the game log files and players' feedback at the end of each game. We refer to the collected log files as game cases and indicate them as CG3.

- Later, subjects P1 to P5, belonging to group G1, were summoned again to play the game with DM activated and using all cases collected so far i.e. CG1, CG2 and CG3.

- Eventually, subjects P16 to P20 of group G4, played first with DM and using all cases CG1, CG2, and CG3 and later without DM.

The different game conditions account for two main factors that might influence user experience: a) the presence/absence of the DM in the game, and b) game knowledge and preferences about previous players through the use of differently sized case libraries.

III. USABILITY STUDY: RESULTS

During interaction, we observed and logged player's actions, responses and reactions to the game. We analyzed the data obtained from both a numerical and a qualitative perspective. From a quantitative analysis view, we were especially interested on the overall score that players assigned to the game both in presence and in absence of the DM. We also were interested in discovering whether and to what extent users prefer the addition of hints (thus ultimately the presence of the DM) during gameplay. The results are summarized in Figure 1 and Figure 2 where they are classified according to the four different subject groups along with the average over these groups. A value of 0 corresponds to the lowest rating, a value of 4 to the highest possible one. Figure 1 clearly shows that players like the game more when they play with the DM enabled. The average score across all players for the game without DM was 2.0 while the average score with DM was 2.25 i.e. an improvement of 12.5% (Wilcoxon, p=0.021, Z=-1.95 one-tailed). The improvements over each group of subjects were: 8.3% for G1, 18.2% for G2, 0% for G3, and 25% for G4.
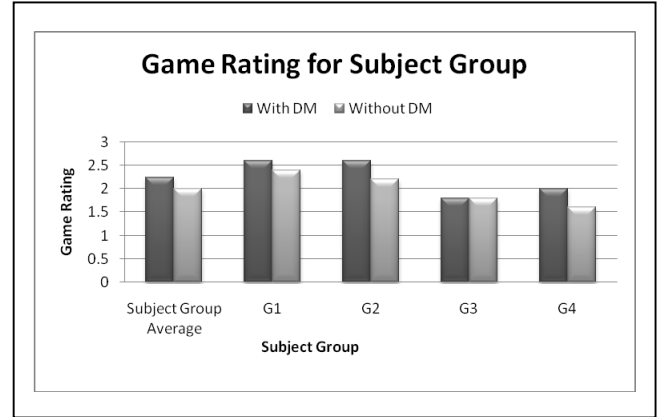

Figure 1. Game ratings according to the different subject groups both with and without DM.

With the collected data we calculated the correlation coefficients to determine the strength and the direction of a linear relationship between several of the variables analyzed. As shown in Figure 3, we calculated the Pearson's product-moment correlation coefficient to relate the degree of experience of players with their game ratings, the degree of experience of players with their hints' ratings, and eventually the game ratings with the hints' rating. The Pearson's coefficient is obtained by dividing the covariance of the two random variables being examined by the product of their standard deviations. Such a degree of linear dependence between two random variables tends to 1 (-1) in case of increasing (decreasing) linear relationships while it assumes some value in between in all other cases. The closer the absolute value of the coefficient is to 1, the stronger the correlation between the two variables. It must be noticed that a value of 0 for the correlation does not necessarily imply that the variables are independent since the Pearson's moment coefficient detects only linear dependencies. The converse is however true i.e. independent variables are characterized by a coefficient 0.

As a consequence, our analysis indicates that people with less game experience tend to rate the game higher as well as to appreciate more game hints. According to our analysis, experienced players seem also to like hints but to a minor degree, which confirms previous studies we carried out [7]. Experienced players and high game rating are instead either not linear correlated or are independent. Pearson's coefficient statistics also reveals that subjects that like the

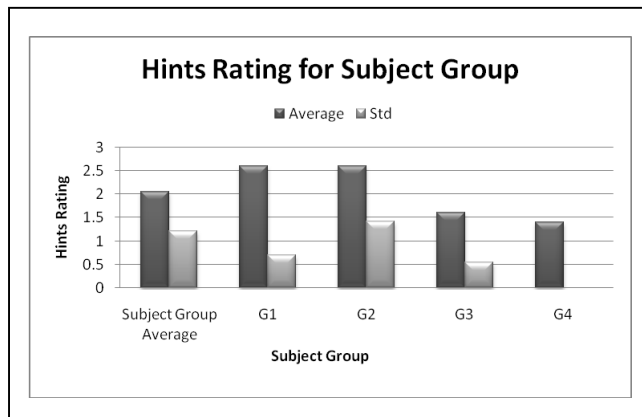game also like to receive hints as they are presented with the DM.

Figure 2.   Average hints' ratings and corresponding standard deviation values for each subject groups.

From the analysis of a post session questionnaire, it turned out that most of the players followed a strategy to play the game (45% did, 35% did sometimes, 15% did not, and 5% don't know) and found it easy to use (25% found it easy, 45% OK, 15% difficult, and 15% very difficult) despite apparently they thought that their input could influence the game only partially.

To the question: "Do you feel you were able to influence the game" 15% said "not at all", 60% "sometimes", and 25% answered "most of the times". 90% of the subjects stated that they would prefer to have speech capabilities enabled along with typed-in input to play the game. The remaining 10% was unable to express or did not have any opinion on that.
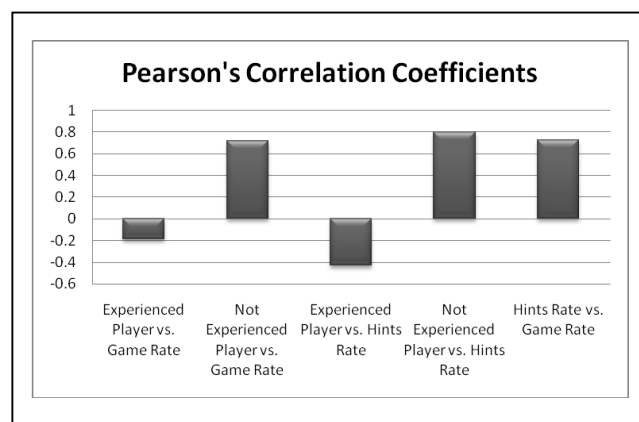

Figure 3.   Pearson's coefficient correlating player experience with hints and game rating.

In order to further understand the interaction, we performed a qualitative analysis using a simplistic version of a well-known qualitative analysis method, Grounded Theory [4]. Grounded Theory is a research technique that operates almost in a reverse fashion to traditional research for it does not begin by researching and developing a hypothesis.

Within such a framework, a variety of data collection methods under different conditions is employed as reported above. Instead of applying a specific model to the phenomenon under investigation, from the data gathered, we identified common patterns (the key points of Grounded Theory), which we then grouped into similar concepts like perception, memory, and language which form the basis for the creation of a reverse engineered hypothesis and which were analyzed singularly.

Here the lessons we learned from such a procedure:

a)   The DM techniques effectively increase player satisfaction. As depicted in Figure 1, a numerical analysis of the interaction already permitted to reach the same conclusion. From the questionnaire data, it emerged that all players perceived a difference between absence and presence of hints despite during their interaction episodes they were never told in which mode they were playing. They were actually not even told that there were different settings between their game episodes.

b)   The type and quantity of DM hints should depend on player's experience. The same conclusion was already reached after numerical analysis of the interaction and highlighted in Figure 2. Moreover, it was found that less experienced players were more likely to be in trouble operating with objects and dealing with situations encountered in the game.

c)   The DM can effectively help in successfully playing the game. Several players pointed out and recognized the utility of the DM to solve riddles in the game. Conversely, players were frustrated when confronted with no hints as they had problems in advancing the game.

d)   The content and quantity of DM hints should be tailored to the player's experience. Currently, the strategy for the presentation of hints during game-play does not take into account the player's personal experience with the game. On that regard, more experienced users complained about system feedback and help messages regarded as not much useful or even trivial. This contradicts with comments of less experienced players who typically reported as mentioned in the previous lessons learned item. Information about the player's previous experience may thus help generating more appropriate hints.

e)   Visual display of game entities must reflect the counterparts (if any) in the real world. This has also to be consistent throughout the game. In fact, a few users reported problems with the perceptions of some game objects. Sometimes, the 2D nature of the game with limited graphical representation made users misrecognize game entities and ultimately issue

commands that were not allowed. Textual descriptions of the graphical environment helped limit this problem.

f) The overt representation of the text presented to display DM hints should reflect the grammar and the vocabulary that can be recognized by the system. From a language point of view, all users quickly understood how to get around invalid commands. Typically, after an average double retyping of a command that did not result in the expected game action, users tend to either try to reformulate the sentence with synonyms or use the words occurring in hints and game descriptions or gave up. After playing for a few minutes, we saw a certain degree of convergence in the users' use of commands. Players tend to make up and retrieve commands from a limited set of commands after finding out that they have worked in a precedent situation. Users just adapted these commands to the situation at hand. The speech capabilities of our system should also be increased to make the game more challenging for demanding players.

g) The type and quantity of DM hints should reflect a player's strategy. The reason for such a conclusion is based on the observation that less experienced users do not follow any particular strategy. They tend to simply follow the hints provided by the DM, and thus remain unsure whether they could actually influence the evolution of the game. They also tend to enjoy more the interaction with non-playing characters since this kind of conversation is currently not focused on the solution of the game. Experienced gamers intelligently pick items that they feel would be useful later in the game and prefer not to waste much time with small-talk with non-playing characters.

h) There should be a balance between game-oriented conversation and small talk when the player interacts with game characters.

IV. CONCLUSIONS

We have presented results from an evaluation study to measure the improvement in user experience with our drama management approach for the interactive fiction game Anchorhead. Our results indicate that quantitatively, the DM causes an average subjective improvement of 12.5% of the subjects' play experience. In particular, a large value of the Pearson's product-moment coefficient highlights a strong correlation between less experienced adventure game players and higher ratings for the inclusion of hints generated by the DM during gameplay. Preference for its inclusion can be evinced also from a qualitative analysis of user interview questionnaires. Moreover, the large majority of players reported that hints are very helpful and also make the game more entertaining. Nonetheless, at times, overexposure to hints that cannot be exploited directly for proceeding through the different situations occurring in the game was considered frustrating. From such an examination, we could draw a series of lessons learned that can help further developing the strategies of the DM.

REFERENCES

[1] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches", AI Communications, 7(1):39-59, 1994.

[2] J. Bates, "Virtual reality art and entertainment", Journal of Teleoperators & Virtual Environments, 2(1):133-138, 1992.

[3] M.S. Gentry, "Anchorhead", Available online at http://www.wurb.com/if/game/17.html, 1998.

[4] B. Glaser, "Basics of grounded theory analysis", Mill Valley, CA: Sociology Press, 1992.

[5] B. Magerko, J. Laird, M. Assanie, A. Kerfoot and D. Stokes, "AI characters and directors for interactive computer games", In: Proceedings of the 2004 Innovative Applications of Artificial Intelligence Conference, 2004.

[6] M. Nelson, M. Mateas, D. Roberts and C. Isbell, "Declarative optimization-based drama management in interactive fiction", IEEE Computer Graphics and Applications, 26(3):33-41, 2006.

[7] M. Sharma, M. Mehta, S. Ontañón and A. Ram, "Player modeling evaluation for interactive fiction", In: Workshop on Optimizing Player Satisfaction at AIIDE 2007, 2007.

[8] P. Weyhrauch, "Guiding Interactive Drama", Ph.D. thesis, Carnegie Mellon University, 1997.

[9] R. Young, M. Riedl, M. Branly, A. Jhala, R. Martin and C. Sagretto, "An architecture for integrating plan-based behavior generation with interactive game environments", Journal of Game Development, 1(1). 2004.

[10] S. Ontañón, A. Jain, M. Mehta and A. Ram, "Developing a drama management architecture for interactive fiction games", In: Proceedings of the 1st Joint International Conference on Interactive Digital Storytelling, 2008.