International Journal on Artificial Intelligence Tools

MODELING AND AUGMENTING GAME ENTERTAINMENT THROUGH CHALLENGE AND CURIOSITY

GEORGIOS N. YANNAKAKIS

Maersk Mc-Kinney Moller Institute
University of Southern Denmark, Campusvej 55
Odense M, 5230, Denmark
georgios@mmmi.sdu.dk

JOHN HALLAM

Maersk Mc-Kinney Moller Institute
University of Southern Denmark, Campusvej 55
Odense M, 5230, Denmark
john@mmmi.sdu.dk

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

This paper presents quantitative measurements/metrics of qualitative entertainment features within computer game environments and proposes artificial intelligence (AI) techniques for optimizing entertainment in such interactive systems. A human-verified metric of interest (i.e. player entertainment in real-time) for predator/prey games and a neuro-evolution on-line learning (i.e. during play) approach have already been reported in the literature to serve this purpose. In this paper, an alternative quantitative approach to entertainment modeling based on psychological studies in the field of computer games is introduced and a comparative study of the two approaches is presented. Feedforward neural networks (NNs) and fuzzy-NNs are used to model player satisfaction (interest) in real-time and investigate quantitatively how the qualitative factors of *challenge* and *curiosity* contribute to human entertainment. We demonstrate that appropriate non-extreme levels of challenge and curiosity generate high values of entertainment and we project the extensibility of the approach to other genres of digital entertainment (e.g. mixed-reality interactive playgrounds).

Keywords: Entertainment modeling; computer games; mixed-reality games; artificial neural networks.

1. Introduction

Computer games, as examples of human-computer interactive systems, provide an ideal environment for research in AI, because they are based on simulations of highly complex and dynamic multi-agent worlds.[1,2,3] Moreover, such systems offer a promising ground for cognitive modeling since they embed rich forms of interactivity between humans and non-player characters (NPCs).[4] Being able to capture quantitatively the level of user (gamer) engagement or satisfaction in real-time can grant insights to the appropriate AI methodology for enhancing the quality of playing experience and furthermore be used to adjust digital entertainment environments according to individual user preferences.[5]

An endeavor on capturing player satisfaction during gameplay and providing quantitative measurements of entertainment in real-time is presented in this article. The principal goal in the reported work is to construct a user model of a class of game playing experience. Specifically, the aim is that the model can predict the answers to which variants of a given game are more or less "fun." This approach is referred to as *Entertainment Modeling*. Herein, entertainment is defined qualitatively primarily as the level of satisfaction generated by the real-time player-game opponent interaction — by 'opponent' we mean any controllable interactive feature of the game. According to this definition, a game is primarily a learning process and the level of entertainment is kept high when game opponents enable new learning patterns ('not too easy a game') for the player that can be perceived and learned by the player ('not too difficult a game').[6,7] On the same basis, according to Kapoor et al. learning is highly correlated to interest, curiosity and intrigue perceived within the axis of emotions varying from boredom to fascination.[8] The collection of these emotions is defined as entertainment (or "fun") in this article.

Entertainment capture in this paper is achieved by following the theoretical principles of Malone's intrinsic qualitative factors for engaging gameplay,[9] namely *challenge* (i.e. *'provide a goal whose attainment is uncertain'*), *curiosity* (i.e. *'what will happen next in the game?'*) and *fantasy* (i.e. *'show or evoke images of physical objects or social situations not actually present'*) and driven by the basic concepts of the theory of *flow* (*'flow is the mental state in which players are so involved in something that nothing else matters'*).[10] Quantitative measures for challenge and curiosity are inspired by previous work on entertainment metrics and extracted from the real-time player-opponent interaction.[6] A mapping between the aforementioned factors and human notion of entertainment is derived using predator/prey games as an initial test-bed.

Two neural network (NN) types, namely a feedforward NN and a fuzzy-NN, are trained through artificial evolution on gameplay experimental data to approximate the function between the examined entertainment factors and player satisfaction. A comparison between the two methods is presented and the methods are validated against and compared with existing metrics of entertainment in the literature.[11] Results demonstrate that both NNs map a function whose qualitative features are consistent with Malone's corresponding entertainment factors and that the evolved feedforward NN provides a more accurate model of player satisfaction for predator/prey games than previous models designed for this genre of games.[6]

The generality of the proposed methodology and results that project its extensibility to other genres of digital entertainment are introduced here. More specifically, the qualitative features of NN mappings between challenge, curiosity and entertainment presented in previous studies[12] and here appear to generalize to games of the mixed-reality Playware interactive game platform.[13] The paper concludes with a discussion of several remaining open questions regarding entertainment modeling and proposes future directions to answer these questions. The limitations of the presented methodology, and the extensibility of the proposed approach of entertainment capture and augmentation are also discussed.

2. Entertainment Modeling

Research in the field of game AI is mainly focused on generating human-like (believable) and intelligent (see Ref. 1, 14 among others) characters. Complex NPC behaviors can emerge through various AI techniques; however, there is no further analysis of whether these behaviors have a positive impact to the satisfaction of the player during play. According to Taatgen et al. [15], believability of computer game opponents, which are generated through cognitive models, is strongly correlated with enjoyable games. Such implicit research hypotheses may well be true; however, there is little evidence that specific NPCs generate enjoyable games unless a notion of interest or enjoyment is explicitly defined.

There have been several psychological studies to identify what is "fun" in a game and what engages people playing computer games. Theoretical approaches include Malone's principles of intrinsic qualitative factors for engaging gameplay,[9] namely challenge, curiosity and fantasy as well as the well-known concepts of the theory of flow [10] incorporated in computer games as a model for evaluating player enjoyment, namely *GameFlow*.[16] A comprehensive review of the literature on qualitative approaches for modeling player enjoyment demonstrates a tendency of overlapping with Malone's and Csikszentmihalyi's foundational concepts. These approaches include Lazzaro's "fun" clustering based on four entertainment factors derived from facial expressions and data obtained from game surveys on players [17]. According to Lazzaro, the four components of entertainment are: hard fun (related to the challenge factor of Malone), easy fun (related to the curiosity factor of Malone), altered states (i.e. 'the way in which perception, behavior, and thought combine in a collective context to produce emotions and other internal sensations' — closely related to Malone's fantasy factor) — and socialization (the people factor). Koster's[7] theory of fun, which is primarily inspired by Lazzaro's four factors, defines "fun" as the act of mastering the game mentally. An alternative approach to fun measure is presented by Read et al.[18] where fun is composed of three dimensions: endurability, engagement and expectations. Questionnaire tools and methodologies are proposed in order to empirically capture the level of fun for evaluating the usability of novel interfaces with children.

Previous work in the field of quantitative entertainment modeling is based on the hypothesis that the player-opponent interaction — rather than the audiovisual features, the context or the genre of the game — is the property that primarily contributes the majority of the quality features of entertainment in a computer game. [6] Based on this fundamental assumption, a metric for measuring the real-time entertainment value of predator/prey games was established as an efficient and reliable entertainment ('interest') metric by validation against human judgement.[19] According to this approach, the three qualitative criteria that collectively define entertainment for any predator/prey game are: the appropriate level of challenge, the opponent behavior diversity and the opponents' spatial diversity. The quantifications of the three criteria provide an estimate — called the I (interest) value, that lies in [0,1] — of real-time entertainment which correlates highly with the human notion of entertainment. Moreover, a real-time learning mechanism based on neuron-evolution is used for augmenting the proposed I value. [19] Similar work in adjusting a game's difficulty includes endeavors through reinforcement learning [20], genetic algorithms [21], probabilistic models [22] and dynamic scripting [23]. However, the aforementioned attempts are based on the assumption that challenge is the only factor that contributes to enjoyable gaming experiences while results reported have not been cross-verified by human players.

As in previous studies,[9,12,6] this paper is primarily focused on the contributions of the opponents' behavior to the entertainment value of the game. However, the work presented here (as an alternative approach to the interest metric[6,19]), rather than being based on empirical observations on human entertainment, presents quantitative measures for Malone's entertainment factors of challenge and curiosity (see also Ref. 12). Given these, the mapping between the two aforementioned factors and the human notion of entertainment, which is based on experimental data from a survey with human players, are extracted (see section 4). Moreover, this article extends the applicability of the AI methodology for entertainment capture introduced by Yannakakis and Hallam.[12] This generalization endeavor is achieved through experiments using the innovative Playware mixed-reality playground.

3. The Test-bed Game

The first test-bed studied is a modified version of the original Pac-Man computer game released by Namco. The player's (*PacMan's*) goal is to eat all the pellets appearing in a maze-shaped stage while avoiding being killed by the four *Ghosts*. The game is over when either all pellets in the stage are eaten by *PacMan*, *Ghosts* manage to kill *PacMan* or a predetermined number of simulation steps is reached without any of the above occurring. In that case, the game restarts from the same initial positions for all five characters. While *PacMan* is controlled by humans, a multi-layered feedforward neural controller is employed to manage the *Ghosts'* motion.

The game is investigated from the opponents' viewpoint and more specifically how the *Ghosts'* adaptive behaviors and the levels of challenge and curiosity they generate can collectively contribute to player satisfaction. The game field (i.e. stage) consists of corridors and walls where both the stage's dimensions and its maze structure are predefined. For the experiments presented in this paper we use a 19×29 grid maze-stage where corridors are 1 grid-cell wide.

We choose predator/prey games as the initial genre of our research in entertainment modeling since, given our aims, they provide us with unique properties. In such games we can deliberately abstract the environment and concentrate on the characters' behavior. Moreover, we are able to easily control a learning process through on-line interaction. Other genres of game (e.g. first person shooters) offer similar properties; however predator/prey games are chosen for their simplicity as far as their development and design are concerned.

4. Experimental Data

The Pac-Man game has been used to acquire data of human judgement on entertainment. To that end, thirty players (13 females, 17 males) whose age covered a range between 17 and 51 years participated in a survey.[19] As part of this survey, each subject plays a set of 25 games against each of two well-behaved opponents (A and B). Each time a pair of sets is finished, the player is asked whether the first set of games was more interesting than the second set of games i.e. whether A or B generated a more interesting game (pairwise preference). In order to minimize any potential of order effects we let each subject play the aforementioned sets in the inverse order too. Statistical analysis on the order effect of game playing on human judgement of entertainment is presented in section 4.1.

As previously mentioned, to capture the subjects' preferences for the game variants played, we use a 2-alternative forced choice (2-AFC) approach since it offers several advantages for a subjective entertainment preference capture: pairwise preference scheme minimizes the assumptions made about subjects' notions of entertainment and allows a fair comparison between the answers of different subjects. Since our focus is to construct a model relating reported entertainment preferences to the levels of entertainment factors that generalizes over the reports of different players, 2-AFC is preferred to a ranking approach.[24] Forcing the choice of subjects generates experimental noise, in that the subject may have no significant preference for one or other of the game variants played yet must nevertheless express a preference; however, insignificant order effects provide evidence that the experimental noise generated in this way is random (see section 4.1).

For the design of the subjects' self reports we follow the principles of comparative fun analysis.[18,25] The endurability and expectations for the majority of subjects that played Pac-Man were very high, indicating that the game design used was successful. More specifically, all subjects were excited to play Pac-Man as soon as they were informed about the rules of the game (derived through a *Funometer* tool application[18]) and the majority of subjects stressed that they would like to play the game again (derived through an Again-Again table[18]).

To validate the I value against human notion of entertainment, subjects in this survey played against five opponents differing in the I value they generate against a well-behaved computer-programmed player in all combinations of pairs.[6] The correlation between human judgement of entertainment and the I value is given by matching the entertainment rankings in which the five opponents are placed by humans and by I value. According to the subjects' answers the I value is correlated highly with human judgement[19] ($r = 0.4444$, p-value = $1.17 \cdot 10^{-8}$). These five opponents are used as a baseline for validating both approaches in this paper (see section 6).

Given the recorded values of human playing times t_k over the 50 (2 times 25) games against a specific opponent, A or B, the average playing time ($E\{t_k\}$) and the standard deviation of playing times ($\sigma\{t_k\}$) for all subjects are computed. We consider the $E\{t_k\}$ and $\sigma\{t_k\}$ values as appropriate measures to represent the level of challenge and the level of curiosity respectively during gameplay[9]. The former provides a notion for a goal whose attainment is uncertain — the lower the $E\{t_k\}$ value, the higher the goal uncertainty and furthermore the higher the challenge — and the latter effectively portrays a notion of unpredictability in the subsequent events of the game — the higher the $\sigma\{t_k\}$ value the higher the opponent unpredictability and therefore the higher the curiosity.

4.1. *Order Effect*

To check whether the order of playing Pac-Man games affects the human judgement of entertainment, we hypothesize that there is no order effect and proceed as follows. For each subject that played a pair of games in both orders, we count the times K and J that the subject chooses the first and the second game respectively as more entertaining in both pairs. In the case where the subject prefers the same opponent in both pairs played, we take no action. The test statistic used to assess the truth of the hypothesis that there is no order effect is given by $z(K, J) = (K - J)/N_s$, where $N_s = 30$. The greater the absolute value of $z(K, J)$ the more the order of play tends to affect the subjects' judgement of entertainment. The obtained z value equals 0.0222 and its corresponding (Trinomial distributed) p-value equals 0.4818. Therefore, the order effect null hypothesis is not rejected and it seems that the order of play does not significantly affect human judgement of entertainment.

5. Tools

Two alternative neural network structures (a feedforward NN and a fuzzy-NN) for learning the relation between the challenge and curiosity factors and the entertainment value of a game have been used and are presented here. The assumption is that the entertainment value y of a given game is an unknown function of $E\{t_k\}$ and $\sigma\{t_k\}$, which the NN will learn. The subjects' expressed preferences constrain but do not specify the values of y for individual games. In other words, we require that comparing the y values assigned by the NN to pairs of games should match human judgement — games judged more entertaining should have higher y. The error is thus determined by the number of pairs ranked incorrectly. Since the output error function is not differentiable, NN training algorithms such as back-propagation are inapplicable. Learning is achieved through artificial evolution [26] and is described in section 5.3.

5.1. *Feedforward NN*

A fully-connected multi-layered feedforward NN has been evolved [26] for the experiments presented here. The sigmoid function is employed at each neuron, the connection weights take values from -5 to 5 and both input values are normalized into [0, 1] before they are entered into the NN. In an attempt to minimize the NN's size, it was determined that single hidden-layered architectures, containing 20 hidden neurons, are capable of successfully obtaining solutions of high fitness (network topology is not evolved, however).

5.2. *Fuzzy-NN*

A fuzzy [27] Sugeno-style [28] inference neural network is trained to develop fuzzy rules by evolving the membership functions for both the input ($E\{t_k\}$, $\sigma\{t_k\}$) and the output variable y of the network as well as each fuzzy rule's weight. Each of the input and output values is represented using five fuzzy sets corresponding to VERY LOW, LOW, AVERAGE, HIGH and VERY HIGH. The membership functions for the input values are triangular and their center α and width β are evolved whereas the output fuzzy sets use singleton membership functions [28] — only the center α of the spike membership function is evolved. The centroid technique is used as a defuzzification method.

5.3. *Genetic Algorithm*

A generational genetic algorithm (GA) [29] is implemented, which uses an "exogenous" evaluation function that promotes the minimization of the difference in matching the human judgement of entertainment. The feedforward NNs and fuzzy-NNs are themselves evolved. In the algorithm presented here, the evolving process is limited to the connection weights of the feedforward NN and the rule weights and membership function parameters of the fuzzy-NN.

The evolutionary procedure used can be described as follows. A population of N networks is initialized randomly. For feedforward NNs, initial real values that lie within [-5, 5] for their connection weights are picked randomly from a uniform distribution, whereas for the fuzzy-NNs, initial rule weight values equal 0.5 and their membership function parameter values lie within [0, 1] (uniformly distributed). Then, at each generation:

Step 1 Each member (neural network) of the population gets two pairs of $(E\{t_k\}, \sigma\{t_k\})$ values one for A and one for B and returns two output values, namely $y_{j,A}$ (output of the game against opponent A) and $y_{j,B}$ (output of the game against opponent B) for each pair j of sets played in the survey ($N_s = 30$). When the $y_{j,A}$, $y_{j,B}$ values are consistent with the judgement of subject j then we state that: 'the values agree with the subject' throughout this paper. In the opposite case, we state that: 'the values disagree with the subject.'

Step 2 Each member i of the population is evaluated via the fitness function f_i:

$$f_i = \sum_{j=1}^{N_s} \begin{cases} 1, & \text{if } y_{j,A}, y_{j,B} \text{ agree with subject } j; \\ \left(\frac{1-|y_{j,A}-y_{j,B}|}{2}\right)^2, & \text{if } y_{j,A}, y_{j,B} \text{ disagree with subject } j. \end{cases} \quad (1)$$

Step 3 A fitness-proportional scheme is used as the selection method.

Step 4 Selected parents clone an equal number of offspring so that the total population reaches N members or reproduce offspring by crossover. The Montana and Davis[30] and the uniform crossover operator is applied for feedforward NNs and fuzzy-NNs respectively with a probability $p_c = 0.4$.

Step 5 Gaussian mutation occurs in each gene (connection weight) of each offspring's genome with a small probability $p_m = 1/n$, where n is the number of genes.

The algorithm is terminated when either a good solution ($f_i > 29.0$) is found or a large number of generations g is completed ($g = 10000$).

6. Results

Results obtained from both feedforward NN and fuzzy-NN evolutionary approaches are presented in this section. In order to control for the non-deterministic effect of the GA initialization phase, each learning procedure (i.e. GA run) for each NN type is repeated ten times — we believe that this number is adequate to illustrate a clear picture of the behavior of each mechanism — with different random initial conditions.

6.1. *Evolving Feedforward NN*

For space considerations, only the two fittest solutions achieved from the evolving feedforward NN approach are illustrated in Fig. 1. The qualitative features of the surfaces plotted in Fig. 1 appeared in all ten learning attempts. The most important conclusions derived from the feedforward NN mapping between $E\{t_k\}$, $\sigma\{t_k\}$ and entertainment are that:

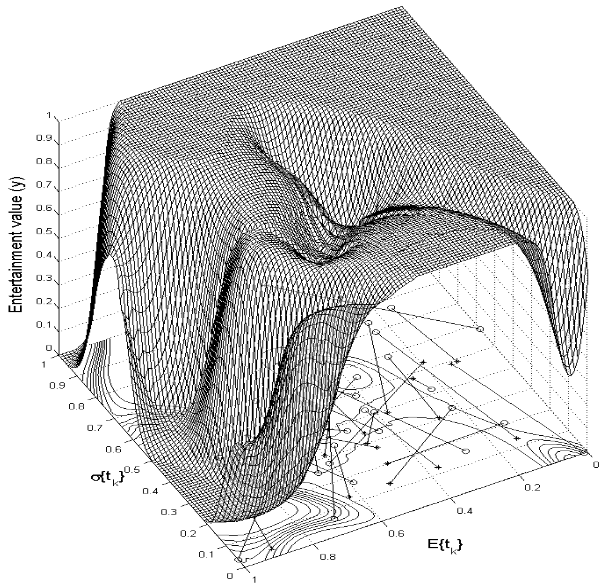
- Entertainment has a low value when challenge is too high ($E\{t_k\} \approx 0$) and curiosity is low ($\sigma\{t_k\} \approx 0$).
- Even if curiosity is low, if challenge is at an appropriate level $(0.2 < E\{t_k\} < 0.6)$, the game's entertainment value is high.
- If challenge is too low $(E\{t_k\} > 0.6)$, the game's entertainment value appears to drop, independently of the level of curiosity.
- There is only a single data point present when $\sigma\{t_k\} > 0.8$ and the generalization of the evolved feedforward NNs within this space appears to be poor. Given that only one out of 60 different gameplay data paints falls in that region of the $E\{t_k\}$-$\sigma\{t_k\}$ two-dimensional space, we can hypothesize that there is low probability for a game to generate curiosity values higher than 0.8. Thus, this region can be safely considered insignificant for these experiments. However, more samples taken from a larger gameplay survey would be required to effectively validate this hypothesis.

The fittest evolved feedforward NN is also tested against the human-designed I metric[6] for cross-validation purposes. The feedforward NN ranks the five different opponents previously mentioned in section 4 in the order $I_1 = I_2 < I_4 < I_3 < I_5$ (where I_i is the entertainment value the i opponent generates) which yields a correlation of 0.5432 (p-value = $3.89 \cdot 10^{-12}$) of agreement with human notion of entertainment expressed by the subject choices in the original experiment. Given this ranking of entertainment against these five opponents, the feedforward NN approach appears to model human entertainment better than the custom-designed interest metric[6] ($r = 0.4444$, p-value = $1.17 \cdot 10^{-8}$).

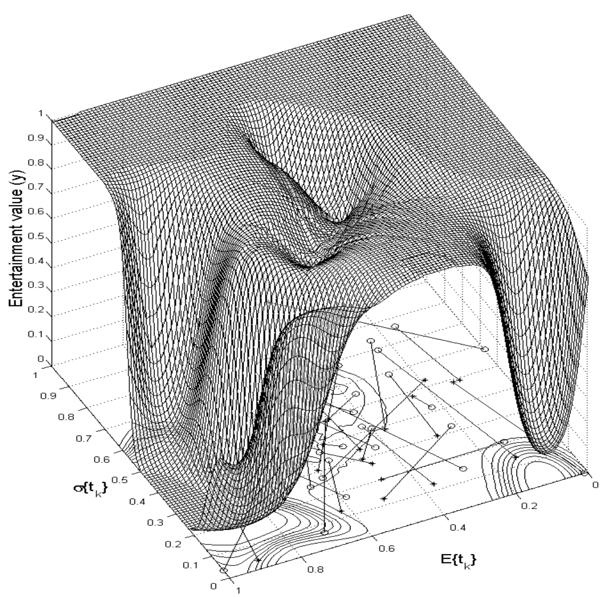
The relationship between entertainment, challenge and curiosity expressed by the evolved feedforward NNs appears to follow the qualitative principles of Malone's work[9] and the human-verified interest metric developed Yannakakis and Hallam[6] for predator/prey games. According to these, a game should maintain an appropriate level of challenge and curiosity in order to be entertaining. In other words, too difficult and/or too easy and/or too unpredictable and/or too predictable opponents to play against make the game uninteresting.

6.2. *Evolving Fuzzy-NN*

The evolutionary procedure for the fuzzy-NN approach is also repeated ten times and only the fuzzy-NN that generates the highest fitness ($f = 29.81$) is presented here for space



(a) The fittest feedforward NN solution ($f = 29.95$).



(b) The second fittest feedforward NN solution ($f = 29.67$).

Fig. 1. Circles ('o') and stars ('$*$') represent $E\{t_k\}$, $\sigma\{t_k\}$ values obtained by playing against opponents A and B respectively. Straight lines are used to connect the sets of games that humans played in pairs.

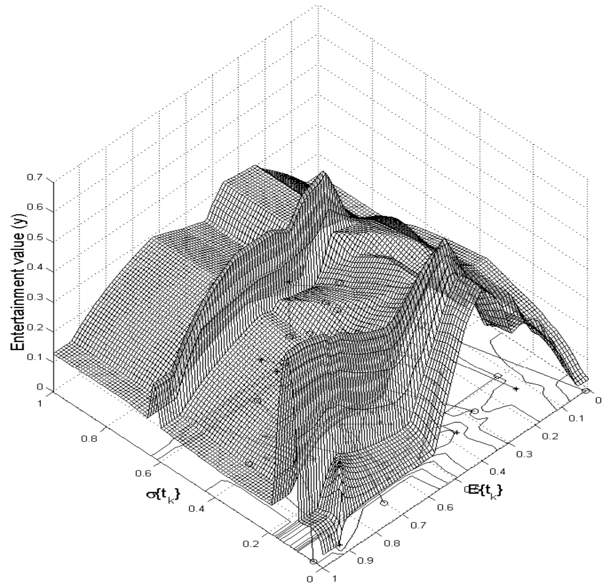


Fig. 2. The fittest fuzzy-NN solution ($f = 29.81$). Circles ('o') and stars ('$*$') represent $E\{t_k\}$, $\sigma\{t_k\}$ values obtained by playing against OLL and NOLL opponents respectively. Straight lines are used to connect the sets of games that humans played in pairs.

considerations. Twenty five fuzzy rules are initially designed based on the conclusions derived from the evolved feedforward NNs. The fittest fuzzy-NN (see Fig. 2) generates 19 fuzzy rules in total — rules with weight values less than 0.1 are not considered significant and therefore are excluded from further consideration — which are presented here with their corresponding weight values w:

- Entertainment is VERY LOW if (a) challenge is VERY HIGH and curiosity is LOW (Rule 1; w_1 = 0.4440) and (b) challenge is LOW and curiosity is AVERAGE (Rule 2; w_2 = 0.3617).
- Entertainment is LOW if (a) challenge is VERY LOW and curiosity is AVERAGE (Rule 3; w_3 = 0.9897) or LOW (Rule 4; w_4 = 0.7068); (b) challenge is LOW and curiosity is HIGH (Rule 5; w_5 = 0.7107); (c) challenge is HIGH and curiosity is VERY LOW (Rule 6; w_6 = 0.5389) and (d) challenge is VERY HIGH and curiosity is VERY LOW (Rule 7; w_7 = 0.9520) or HIGH (Rule 8; w_8 = 0.9449).
- Entertainment is AVERAGE if challenge is VERY LOW and curiosity is HIGH (Rule 9; w_9 = 0.5818).
- Entertainment is HIGH if (a) challenge is LOW and curiosity is VERY LOW (Rule 10; w_{10} = 0.8498) or VERY HIGH (Rule 11; w_{11} = 0.2058); (b) challenge is AVERAGE and curiosity is LOW (Rule 12; w_{12} = 0.5); (c) challenge is HIGH and curiosity is LOW (Rule 13; w_{13} = 0.2824) or AVERAGE (Rule 14; w_{14} = 0.25) and (d) challenge is VERY HIGH and curiosity is AVERAGE (Rule 15; $w_1 5$ = 0.2103).
- Entertainment is VERY HIGH if (a) challenge is VERY LOW and curiosity is VERY HIGH

(Rule 16; w_{16} = 0.7386); (b) challenge is AVERAGE and curiosity is VERY LOW (Rule 17; w_{17} = 0.5571) or VERY HIGH (Rule 18; w_18 = 0.8364) and (c) challenge is HIGH and curiosity is HIGH (Rule 19; w_{19} = 0.2500).

The majority of the fuzzy rules generated by the neuro-fuzzy approach appear consistent with Malone's principles on challenge and curiosity, the empirical contributions of the interest metric from the literature[19] and the fittest feedforward NN presented in section 6.1. However, the fittest fuzzy-NN (being less fit than the fittest feedforward NN) generates some few fuzzy rules that are not consistent with the aforementioned principles — e.g. Rule 10: entertainment is HIGH if challenge is LOW and curiosity is VERY LOW. It is not clear whether the poorer performance is intrinsic to the method or a result of unlucky initialization; further tests are needed to distinguish these alternatives.

The fuzzy-NN is tested against the I metric of section 4 as in the evolved feedforward NN approach. The evolved fuzzy-NN ranks the five opponents in the order $I_2 < I_1 < I_3 < I_4 = I_5$. This ranking demonstrates a correlation of 0.3870 (p-value = $1.74006 \cdot 10^{-6}$) of agreement with human notion of entertainment, which appears to be lower than the correlation achieved through the I value of section 4 ($r = 0.4444$, p-value = $1.17 \cdot 10^{-8}$). However, as in the feedforward NN approach, the generalization of the evolved fuzzy-NNs appears to be poor when $\sigma\{t_k\} > 0.8$ due to the presence of a single data point within this region of the $(E\{t_k\}, \sigma\{t_k\})$ space. Even though we consider this non-frequent region insignificant as far as this work is concerned, it may be sampled from a more extensive human game experiment in a future study.

7. Extensibility of the Approach

This section examines the hypothesis that results obtained from experiments with predator/prey computer games can scale to different genres of game. The experiments on the Playware playground presented here offer a comparison with games of a very different genre. Playware, being a mixed-reality interactive playground designed for children, extends our hypothesis testing to dissimilar game platforms and age groups of players. The section includes a brief presentation of the Playware playground and the test-bed used and concludes with an analysis of the entertainment models obtained by following the same experimental procedure as in the Pac-Man game. A feedforward NN is used here, since that was the most accurate model of human notion of entertainment found for the Pac-Man game.

7.1. *Playware Playground*

New emerging playing technologies (e.g. computer games) have contributed to transforming the way children spend their leisure time: from outdoor or street play to play sitting in front of a screen. This sedentary style of play may have health implications.[31] A new generation of playgrounds that adopt technology met in computer games may address this issue. More specifically, intelligent interactive playgrounds with abilities of adapting the game according to each child's personal preferences provide properties that can keep children engaged in entertaining physical activity. On that basis, capturing the child's entertainment and adjusting the game in order to increase it can only have positive effects on the child's physical condition. The Playware playground adopts these primary concepts.

The Playware[32] prototype playground consists of several building blocks (i.e. tangible tiles — see Fig. 3) that allow for the game designer to develop a significant number of different games within the same platform. For instance, tiles can be placed on the floor or on the wall in different topologies to create a new game.[32] The overall technological concept of Playware is based on embodied AI[33] where intelligent physical identities (tiles) incorporate processing power, communication, input and output, focusing on the role of the morphology-intelligence interplay in developing game platforms.

7.1.1. *Bug-Smasher Game*

The test-bed game used for the experiments presented here is called 'Bug-Smasher'. The game is developed on a 6 x 6 square tile topology (see Fig. 3). During the game, different 'bugs' (colored lights) appear on the game surface and disappear sequentially after a short period of time. A bug's position is picked within a radius of three tiles from the previous bug and according to a predefined level of the bugs' spatial diversity (see Ref. 34 for more details). The child's goal is to smash as many bugs as possible by stepping on the lighted tiles. Different sounds and colors represent different bugs when appearing and when smashed in order to increase the fantasy entertainment factor.[9] Bug-smasher is an action physical game that enhances children's pattern recognition and reflective skills. In that sense, it is related to the genre of predator/prey games previously studied — see Ref. 34 for a detailed description of Bug Smasher.

7.2. *Experimental Data*

The Bug-Smasher game has also been used to acquire data of human judgement on entertainment. We consider the speed (S — in sec^{-1}) that the bugs appear and disappear from the game and their spatial diversity (measured by bugs' tile visit entropy H) on the game's plane as appropriate measures to represent the level of challenge and the level of curiosity (unpredictability) respectively [9] during gameplay. As in the Pac-Man game, the former provides a notion for a goal whose attainment is uncertain (challenge) and the latter effectively portrays a notion of unpredictability in the subsequent events of the game (curiosity).

28 children whose age covers a range between 8 and 10 years participated in an experiment following the same comparative fun procedure as described in section 4. Statistical analysis (following the procedure presented in section 4.1) of the subjects' answers shows that the order effect on children judgement on entertainment is not statistically significant ($r_c = -0.0714$, p-value$= 0.3444$). As in the Pac-Man game, the levels of challenge and curiosity and the subjects' answers are used to guide the training of a feedforward NN model of entertainment using the methods described in sections 5.3 and 6.1.

Fig. 3. A child playing the Bug-Smasher game.

7.3. *Results*

The experiment presented here tests the hypothesis of the existence of a generic notion of entertainment given the level of challenge and curiosity in dissimilar games. Given the 30 pairs of games, where the games have different levels of S and/or H, a feedforward NN is evolved by following the approach presented in section 5.3. Experiments with NN topologies consisting of (maximum) two hidden layers with up to 30 hidden neurons showed that a single hidden-layered architecture, containing 10 hidden neurons is the minimal NN architecture capable of successfully obtaining solutions of high fitness.

The fittest NN found was able to correctly match 20 out of 30 children answers on entertainment. Such a poor fitness indicates the difficulty of adjusting values of challenge and curiosity for inferring entertainment values in an objective manner (without the presence of individual playing characteristics). In contrast with the metrics of challenge ($E\{t_k\}$) and curiosity ($\sigma\{t_k\}$) derived for the Pac-Man game, bug speed (S) and bug spatial diversity (H) do not embed any feature of player-opponent interaction since they define opponent/bug (controllable game feature) characteristics. Therefore, the relation between challenge (S), curiosity (H) and the game's entertainment value (y) illustrated in Fig. 4 projects a rather objective enjoyment model for the Bug-Smasher game.

Despite the best solution's poor fitness, the correlation between entertainment, challenge and curiosity generated through the evolved NN (see Fig. 4) appears to be consistent with the entertainment model obtained through Pac-Man and Malone's qualitative principles.[9] A conclusion derived from experiments in both games is that extreme levels of challenge and curiosity make the game uninteresting. As seen from Fig. 4, average levels of challenge ($0.5 < S < 0.8$) and curiosity ($0.3 < H < 0.9$) generate high entertainment values. Moreover, it appears that games of the lowest challenge level ($S \approx 0$) combined with the highest curiosity level ($H \approx 1$) may yield high entertainment values. The entertainment model presented here, being consistent with features of the Pac-Man entertainment models

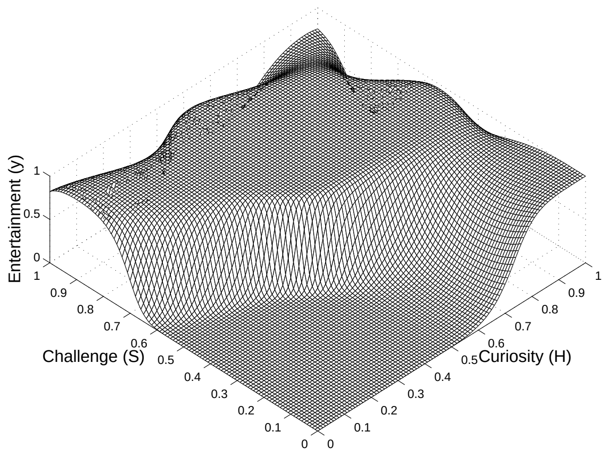


Fig. 4. Fittest feedforward NN ($f = 22.82$).

(see Fig. 1 and Fig. 2), demonstrates that the proposed approach generalizes to dissimilar games (Pac-Man vs. Bug-Smasher), game platforms (computer vs. mixed-reality Playware) and age groups (adults vs. children).

8. Conclusions

This paper introduced an approach to constructing a quantitative value of entertainment motivated by the qualitative principles of Malone's intrinsic factors for engaging game play.[9] More specifically, the quantitative impact of the factors of challenge and curiosity on human entertainment were investigated in computer (Pac-Man) and mixed-reality (Bug-Smasher) games.

The two neuro-evolution approaches for modeling entertainment examined in the Pac-Man game demonstrate qualitative features that share principles with the human-verified interest metric (I value) reported in the literature.[6] Both approaches manage to map successfully between the measures of entertainment factors such as challenge and curiosity and the notion of human game play satisfaction. The proposed approaches replace the hand-crafted mathematical formulation of the interest metric with a more general process of machine learning applied to neural network models.

Validation results obtained show that the fittest feedforward NN gets closer to human notion of entertainment than both the I value[19] and the fittest fuzzy-NN. Therefore, it appears that the average and the standard deviation of a human's playing time over a number of games are sufficient in themselves and more effective than the I value (as reported in Ref. 19) in capturing player entertainment in real-time in predator/prey games.

The work reported for this approach is most significantly limited by the number of participants in the game survey we devised. Therefore, not all regions of the challenge-curiosity search space were sampled by human play which therefore yielded poor NN generalization for these regions. Limited data also restricted the sensible number of inputs to the learning system.

Malone's entertainment factor of fantasy is omitted here since the focus is on the contribution of the opponent behaviors to the generation of entertainment; however, experiments on interactive physical predator/prey games with children have shown that entertainment increases monotonically with respect to the fantasy factor.[25]

The entertainment modeling approach presented here demonstrates generality over dissimilar games through experiments in the interactive mixed-reality Playware playground. The successful (in the Pac-Man game) feedforward NN was used in this case to model entertainment. Results obtained provided evidence that the generated models, which map between challenge curiosity and entertainment, share the same qualitative features independently of game genre, game platform and age group. These features match Malone's principles for engaging gameplay: non-extreme levels of challenge and curiosity generate high levels of entertainment. This consistency shows that neuro-evolution based on comparative fun experiments constitutes a generic approach for modeling entertainment effectively in games.

9. Discussion

The foregoing has proposed and demonstrated methods for deriving a quantitative estimate of the level of entertainment experienced by a player of a interactive entertainment system (computer and mixed-reality game), using data that can be derived from or during game play. In this section, we discuss some of the questions raised by the approach and the assumptions on which it is based.

An immediate and natural question is whether the techniques described really capture "entertainment" which, after all, is a complex mental phenomenon depending on the player, the game and (probably) a number of external factors in rather involved ways. We acknowledge that the definition of a quantitative metric for entertainment in this sense is almost certainly infeasible. However, we take a practical approach here: it is sufficient for our purposes if a quantity exists that can be computed from observable data from the player-game interaction and that correlates well with players' expressed preferences. In other words, a numerical value which orders games in the same way as players' judgement of entertainment is sufficient for our purposes.

The foregoing material illustrates a way to construct such a metric: using machine learning to explore the space of possible evaluation functions whose values are consistent with human judgement of entertainment value. The resulting metric is specific to the particular game under consideration, but the general method of construction is applicable to a wide variety of game instances and genres.

To summarize, therefore, the proposed approach does not capture details of the complex mental states associated with enjoyment of computer or mixed-reality games, but it does provide a way to evaluate different instances of game play in terms of how entertaining they are for a player. Such knowledge can be used, for example, for tuning the game to suit the player (see below).

In addition to this general question concerning the approach, there are a number of assumptions (and hence limitations) associated with the methods presented; these are discussed below.

9.1. *Assumptions and Limitations of the Proposed Approach*

The method proposed is based on the same fundamental assumptions as the I value approach: that the opponents' behavior is the primary determinant of the entertainment value of a given instance of game play. There are two limitations specific to this methodology described as follows.

- The effectiveness of a machine learning method depends strongly on the quantity and quality of data available. For the cases considered here, this data comprises two kinds: raw measurements derived from game play, that represent aspects of features such as challenge, curiosity, fantasy, etc.; and expressed preferences or rankings between instances of game play. The former provide the observables on which the metric is built, the latter determine the degree of consistency with human judgement of any given proposal for the metric function.

  Obtaining the latter kind of data involves experimentation in which players are asked to say which of several (here two) instances of game play they prefer; collecting such data is time- and player- consuming. This limits the complexity of metric that can be considered, since it limits the feedback available to the machine learning process during the exploration of the space of metrics.
- The former kind of data also presents certain problems: specifically, how do we know what measurements to include as a basis for the metric? The work presented here uses two different pairs of designer-chosen measurement for each relevant feature — challenge and curiosity — and game — Pac-Man and Bug-Smasher — but one could in principle devise many measurements correlated with either feature, and allow the machine learning system to use all of them or to make a selection of the best measurements for the purpose. This approach removes a certain designer bias in the method at the expense of complicating the machine learning task and approaching earlier the limits imposed by the difficulty of collecting preference data from players.

9.2. *Making Use of Entertainment Models*

Given a successful metric of entertainment for a given game, designed and evaluated using one of the methods proposed above, the final question we consider here is how such knowledge might be used. As previously noted, opponents which can learn and adapt to new playing strategies offer a richer interaction to entertain the player. An obvious use of the entertainment metric is therefore to adapt the game so that the metric value increases.

A possible strategy for this might be to use a metric evaluation function constructed using the machine learning technique directly to enhance the entertainment provided by the game. The key to this is the observation that the models (feedforward NN or fuzzy-NN) relate game features to entertainment value. It is therefore possible in principle to infer what changes to game features will cause an increase in the interestingness of the game, and to adjust game parameters to make those changes. For the feedforward NN, the partial derivatives of $\vartheta y/\vartheta E\{t_k\}$ and $\vartheta y/\vartheta \sigma\{t_k\}$ for the Pac-Man game and $\vartheta y/\vartheta S$ and $\vartheta y/\vartheta \sigma H$ for the Bug-Smasher game indicate the change in entertainment for a small change in an individual game feature. One could use gradient ascent to attempt to improve entertainment with such a model. The fuzzy-NN approach provides qualitative rules relating game features to entertainment, rather than a quantitative function, but an analogous process could be applied to augment game entertainment.

Such a direction constitutes an example of future work within computer and physical games. The level of engagement or motivation of the user/player/gamer of such interactive environments can be increased by the use of the presented approaches. Apart from providing systems of richer interaction and qualitative entertainment,[5] such approaches can generate augmented motivation of the user for deep learning in learning environments that use games (i.e. edutainment).

Acknowledgements

This work was supported in part by the Danish Research Agency, Ministry of Science, Technology and Innovation (project no: 274-05-0511).

References

1. John E. Laird and Michael van Lent. Human-level AI's killer application: Interactive computer games. In *Proceedings of the Seventh National Conference on Artificial Intelligence (AAAI)*, pages 1171–1178, 2000.
2. John D. Funge. *Artificial Intelligence for Computer Games*. A. K. Peters Ltd, Wellesley, Massachusetts, USA.
3. Alex J. Champandard. *AI Game Development*. New Riders Publishing, 2004.
4. Peter Vorderer, Tilo Hartmann, and Cristoph Klimmt. Explaining the enjoyment of playing video games: the role of competition. In D. Marinelli, editor, *ICEC conference proceedings 2003: Essays on the future of interactive entertainment*, pages 107–120, Pittsburgh. Carnegie Mellon University Press.
5. Georgios N. Yannakakis and John Hallam. A scheme for creating digital entertainment with substance. In *Proceedings of the Workshop on Reasoning, Representation, and Learning in Computer Games, 19th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 119–124, August 2005.
6. Georgios N. Yannakakis and John Hallam. Evolving Opponents for Interesting Interactive Computer Games. In S. Schaal, A. Ijspeert, A. Billard, Sethu Vijayakumar, J. Hallam, and J.-A. Meyer, editors, *From Animals to Animats 8: Proceedings of the 8^{th} International Conference on Simulation of Adaptive Behavior (SAB-04)*, pages 499–508, Santa Monica, LA, CA, July 2004. The MIT Press.
7. Ralph Koster. *A Theory of Fun for Game Design*. Paraglyph Press, 2005.
8. Ashish Kapoor, Selene Mota, and Rosalind Picard. Towards a Learning Companion that Recognizes Affect. In *Proceedings of Emotional and Intelligent II: The Tangled Knot of Social Cognition, AAAI Fall Symposium* , November 2001.
9. Thomas W. Malone. What makes computer games fun? *Byte*, 6:258–277, 1981.
10. M. Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*. New York: Harper & Row, 1990.

11. Georgios N. Yannakakis and John Hallam. A Generic Approach for Obtaining Higher Entertainment in Predator/Prey Computer Games. *Journal of Game Development*, 1(3):23–50, December 2005.
12. Georgios N. Yannakakis and John Hallam. Towards Capturing and Enhancing Entertainment in Computer Games. In *Proceedings of the 4^{th} Hellenic Conference on Artificial Intelligence, Lecture Notes in Artificial Intelligence*, volume 3955, pages 432–442, Heraklion, Greece, May 2006. Springer-Verlag.
13. Henrik Hautop Lund and C. Jessen. Playware — intelligent technology for children's play. Technical Report TR-2005-1, Maersk Institute For Production Technology, University of Southern Denmark, June 2005.
14. Alexander Nareyek. Intelligent agents for computer games. In T.A. Marsland and I. Frank, editors, *Computers and Games, Second International Conference, CG 2002*, pages 414–422, 2002.
15. N. A. Taatgen, M. van Oploo, J. Braaksma, and J. Niemantsverdriet. How to construct a believable opponent using cognitive modeling in the game of set. In *Proceedings of the fifth international conference on cognitive modeling*, pages 201–206, 2003.
16. Penelope Sweetser and Peta Wyeth. GameFlow: A Model for Evaluating Player Enjoyment in Games. *ACM Computers in Entertainment*, 3(3), July 2005.
17. Nicole Lazzaro. Why we play games: Four keys to more emotion without story. Technical report, XEO Design Inc., 2004.
18. Janet Read, Stuart MacFarlane, and Chris Cassey. Endurability, engagement and expectations. In *Proceedings of International Conference for Interaction Design and Children*, 2002.
19. Georgios N. Yannakakis and John Hallam. Towards Optimizing Entertainment in Computer Games. *Applied Artificial Intelligence*, 2007. to appear.
20. Gustavo Andrade, Geber Ramalho, Hugo Santana, and Vincent Corruble. Extending reinforcement learning to provide dynamic game balancing. In *Proceedings of the Workshop on Reasoning, Representation, and Learning in Computer Games, 19th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 7–12, August 2005.
21. Milan A. Verma and Peter W. McOwan. An adaptive methodology for synthesising Mobile Phone Games using Genetic Algorithms. In *Congress on Evolutionary Computation (CEC-05)*, pages 528–535, Edinburgh, UK, September 2005.
22. Robin Hunicke and Vernell Chapman. AI for Dynamic Difficulty Adjustment in Games. In *Proceedings of the Challenges in Game AI Workshop, 19^{th} Nineteenth National Conference on Artificial Intelligence (AAAI'04)*, 2004.
23. Pieter Spronck, Ida Sprinkhuizen-Kuyper, and Eric Postma. Difficulty Scaling of Game AI. In *Proceedings of the 5th International Conference on Intelligent Games and Simulation (GAME-ON 2004)*, pages 33–37, 2004.
24. R. L. Mandryk, K. M. Inkpen, and T. W. Calvert. Using Psychophysiological Techniques to Measure User Experience with Entertainment Technologies. *Behaviour and Information Technology (Special Issue on User Experience)*, 25(2):141–158, 2006.
25. Georgios N. Yannakakis, John Hallam, and Henrik Hautop Lund. Comparative Fun Analysis in the Innovative Playware Game Platform. In *Proceedings of the 1^{st} World Conference for Fun 'n Games*, pages 64–70, 2006.
26. X. Yao. Evolving artificial neural networks. In *Proceedings of the IEEE*, volume 87, pages 1423–1447, 1999.
27. L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.
28. M. Sugeno. *Indstrial Applicatios of Fuzzy Control*. North-Holland, 1985.
29. John H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, 1975.
30. D. J. Montana and L. D. Davis. Training feedforward neural networks using genetic algorithms. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-*

89), pages 762–767, San Mateo, CA, 1989. Morgan Kauffman.
31. M. Lindstrom and P.B. Seybold. *BRANDchild: Insights into the Minds of Today's Global Kids: Understanding Their Relationship with Brands*. Kogan Page, 1994.
32. Henrik Hautop Lund, Thomas Klitbo, and C. Jessen. Playware technology for physically activating play. *Artifical Life and Robotics Journal*, 9(4):165–174, 2005.
33. R. Pfeifer and C. Scheier. *Understanding Intelligence*. Cambridge, MIT Press, 1999.
34. Georgios N. Yannakakis, Henrik Hautop Lund, and John Hallam. Modeling Children's Entertainment in the Playware Playground. In *Proceedings of the IEEE Symposium on Computational Intelligence and Games*, pages 134–141, Reno, USA, May 2006. IEEE.