

Preference Learning for Cognitive Modeling: A Case Study on Entertainment Preferences

Georgios N. Yannakakis, *Member, IEEE*, Manolis Maragoudakis, *Member, IEEE*, and John Hallam

Abstract—Learning from preferences, which provide means for expressing a subject’s desires, constitutes an important topic in machine learning research. This paper presents a comparative study of four alternative instance preference learning algorithms (both linear and nonlinear). The case study investigated is to learn to predict the expressed entertainment preferences of children when playing physical games built on their personalized playing features (*entertainment modeling*). Two of the approaches are derived from the literature—the large-margin algorithm (LMA) and preference learning with Gaussian processes—while the remaining two are custom-designed approaches for the problem under investigation: meta-LMA and neuroevolution. Preference learning techniques are combined with feature set selection methods permitting the construction of effective preference models, given suitable individual playing features. The underlying preference model that best reflects children preferences is obtained through neuroevolution: 82.22% of cross-validation accuracy in predicting reported entertainment in the main set of game survey experimentation. The model is able to correctly match expressed preferences in 66.66% of cases on previously unseen data (p -value = 0.0136) of a second physical activity control experiment. Results indicate the benefit of the use of neuroevolution and sequential forward selection for the investigated complex case study of cognitive modeling in physical games.

Index Terms—Augmented-reality games, Bayesian learning (BL), entertainment modeling, large-margin classifiers, neuroevolution, preference learning.

I. INTRODUCTION

THERE is an increasing trend toward personalization of services in computer games, e-commerce, marketing, as well as other fields. The use of computational methods for unveiling hidden knowledge regarding the prediction of certain preferences is also of great importance. Preference learning is a research area that aids in the process of exploiting a set of specific features of a subject (e.g., an individual) in an attempt to predict his or her preferences.

Consider, for example, a banker who knows that a middle-aged working man with two children prefers investment A over B over C from an imaginary set of investment programs that the

Manuscript received February 8, 2008; revised September 1, 2008. First published September 29, 2009; current version published October 16, 2009. This work was supported in part by the Danish Research Agency, Ministry of Science, Technology and Innovation, under Project 274-05-0511. This paper was recommended by Associate Editor L. Rothrock.

G. N. Yannakakis is with the IT University of Copenhagen, 2300 Copenhagen, Denmark (e-mail: yannakakis@itu.dk).

M. Maragoudakis is with the Department of Information and Communication Systems Engineering, University of the Aegean, 83200 Karlovassi, Greece (e-mail: mmarag@aegean.gr).

J. Hallam is with the Maersk Mc-Kinney Moller Institute, University of Southern Denmark, 5230 Odense, Denmark (e-mail: john@mmmi.sdu.dk).

Digital Object Identifier 10.1109/TSMCA.2009.2028152

bank offers. The banker has used previous experience to learn certain preferences. A similar approach can be followed by a machine learner. The aforementioned example can be viewed as an enhanced variation of supervised learning, where instances are tagged with preference relations over a set of possible categorizations, instead of with a single value. From a theoretical perspective, such a situation goes beyond prediction of single values (either in regression or pattern recognition) since it involves construction of preference models, using relational structures or value functions [1].

The context of the work reported here is the construction of user models for a class of children’s gameplaying experience during physical play in the Playware playground platform. The purpose of the model is to predict children’s answers to which variants of a given game are more or less “fun” (or “entertaining,” which is used synonymously in this paper). This cognitive modeling task is referred to as *entertainment modeling*. The word “fun” is used extensively hereinafter since it captures best, in our view, children’s notion of the term “entertainment” [2] and is the term used by the children when making their experimental self-reports. Entertainment generated by a physical-game experience is captured through features extracted from user–game interaction, and feature selection is used for choosing an appropriate set of features that successfully predict expressed entertainment preferences.

In our previous work, we used neuroevolution to construct such models. In this paper, however, four preference learning techniques (large-margin algorithm (LMA) [3], meta-LMA, preference learning with Gaussian processes [4], and neuroevolution) are tested in this application, and their performances are compared. In each case, the output of the constructed model is a real number y such that more enjoyable games receive higher numerical output and that the model functions as an efficient predictor of reported entertainment preferences, given suitable specific individual gameplay features. Suitable input-feature subsets are constructed using two alternative feature selection schemes [n best feature selection (nBest) and sequential forward selection (SFS)], the performances of which are also compared. This basic approach to entertainment modeling is applicable to a variety of games, both computer [5] and physical [6]–[8], using features derived from the interaction of player and game environment measured through game parameters and/or physiological signal data.

The comparative studies between the two feature selection methods and the four preference learning approaches reveal that evolving artificial neural network (ANN) models combined with SFS generate the highest accuracy in classifying between preferred and nonpreferred Playware game variants. These

models are trained and validated on gameplay data obtained from a first (main) set of experimentations and are then evaluated using unseen data from a second gameplay and control-experiment set. The results indicate that ANN user models that are able to predict children's preferred game variants, given suitable individual playing-feature representations, can indeed be constructed and that such models not only distinguish gameplay from gamelike nonentertaining physical activity but also generalize (to some extent) over children's individual preferences.

This paper concludes with a discussion of the limitations of the proposed methodology and of the extent to which it could be applied to other user-modeling problems. Its generic use as an efficient baseline for capturing reported entertainment in physical interactive games in real time is also outlined.

II. PREFERENCE LEARNING

Preference learning has recently attracted significant attention in machine learning research [1], [9]. The problem of preference learning has been viewed in two main ways: *instance* and *label* preference learning [1]. We focus on the former here. Within instance preference learning (IPL), the problem consists of a set of instances \mathcal{X} 's that are associated with a total- or partial-order relation and a set of pairwise preference relations \mathcal{D} 's on the instances. In contrast to supervised learning, training instances are not specified by target values but rather by pairwise preferences between them. The instances are typically characterized by a set of attribute-value features. The goal is to learn a relation among the instances that satisfies those pairwise preferences. Specifically, IPL, in its general form, learns a preference relation $\mathcal{P}_x \subseteq \mathcal{X} \times \mathcal{X}$ for any instance $x \in \mathcal{X}$, given the pairwise preferences $\mathcal{D} = \{v_k \succ u_k | k = 1, \dots, m\}$. Here, $v_k \succ u_k$ means that the instance v_k is preferred to u_k (we assume in the following that v 's are always preferred to u 's). Previous work on this area has presented approaches that learn a utility function that assigns a score to each instance [10], [11]. Moreover, algorithms based on large-margin classifiers (see [3], [10], and [12], among others) and Gaussian processes [4] for preference learning have been proposed and applied to a series of dissimilar IPL case studies.

For the entertainment modeling problem under investigation, the mechanism attempts to approximate a function that predicts whether $u_1 \succ u_2$ holds, where u_1 and u_2 represent particular instances of individual gameplay, for any $u_1, u_2 \in \mathcal{X} \subset \mathcal{R}^d$, given the following:

- 1) a set of d features (features extracted from player–game interaction data) characterizing a gameplay instance;
- 2) a set of n training instances $\mathcal{X} = \{x_i | i = 1, \dots, n\}$ comprising vectors of the measured values of those features for the (various) games played by the (various) players;
- 3) a set of m pairwise preferences $\mathcal{D} = \{v_k \succ u_k | k = 1, \dots, m\}$ in which players record which of two games (game variants) that they preferred ($u_k, v_k \in \mathcal{X}$ are the feature value vectors of the specific gameplay instances concerned).

Four different IPL algorithms are used and compared in this paper: LMA [3], an algorithm based on LMA but using

a nonlinear utility function (meta-LMA), IPL with Gaussian processes [4], and evolving feedforward neural networks [13]. (Further details of the mechanisms used can be found in Section III.)

To the best of our knowledge, the work of Fiechter and Rogers [3] is the only comparative study of preference learning approaches including neuroevolution [13]. In contrast to that study demonstrating the advantages of the use of large-margin mechanisms [3] over neuroevolution, in our case study, the inverse occurs: Neuroevolution appears to be the most appropriate IPL mechanism for effectively capturing children's level of entertainment in physical play. While preference learning with neuroevolution has been introduced in the authors' earlier work [6], this paper is innovative in two ways. First, it presents a comparative study of various preference learning mechanisms (two derived from the literature and two designed for the problem under investigation). Second, even though learning from pairwise preferences has been successfully demonstrated in a number of domains like scheduling [14], [15] and search-engine optimization [11], this paper introduces preference learning as a tool for a cognitive modeling application and demonstrates the supremacy of neuroevolution for the problem chosen.

III. PREFERENCE LEARNING MECHANISMS

The proposed approach to entertainment modeling is based on selecting a minimal subset of individual features, on the basis of which is constructed a quantitative user model that predicts the subject's reported entertainment preferences. The assumption is that the entertainment value y of a given game, which models the subject's internal response to playing the game, i.e., how much "fun" it is, is an unknown function of individual features that a machine learning mechanism can learn. The subject's expressed preferences constrain but do not specify the values of y for individual games, but we assume that the subject's expressed preferences are consistent.

Constraint-satisfaction algorithms cannot solve the problem since the variable y under the constraint $y_A > y_B$ for any two given games A and B has no specific domain values. Likewise, any machine learning that is based on learning a target output is inapplicable since target outputs are unknown. By the use of a ranking approach, numerical values for the y variable could be made available; however, while ranking is (presumably) internally consistent for a given player, rank between players may not be commensurate, giving spurious numerical targets for learning. Preference learning [9] is the only applicable type of machine learning for this constrained-classification problem. There are several techniques that learn from a set of pairwise preferences, such as algorithms based on support vector machines (SVMs) [3], Gaussian processes [4], and evolving ANNs. The four preference learning mechanisms used and compared in this paper are presented in the following sections.

A. LMA

The LMA introduced in [3] is based on the fundamental theory of SVMs and constitutes the baseline linear preference-learning approach to our problem. LMA has been successfully

applied in routing problems where, among many approaches, it even outperforms evolving ANNs [3]. This algorithm restricts its investigation to subjective entertainment preference functions (y 's), which are linear combinations of individual features f 's, i.e., $y(f) = f \cdot w$. The vector $w = (w_1, w_2, \dots, w_d)$ represents positive weight variables for the d features investigated (i.e., a linear classifier) and optimized by the algorithm.

The goal is to obtain $y(v_k) > y(u_k)$ since the child prefers the gameplay instance with feature vector v_k over one with feature vector u_k for each pairwise preference comparison k . Let $f_k = u_k - v_k$, $k = 1, \dots, m$; the classifier with large margin can then be obtained by solving the following linear programming problem using (for example) the simplex algorithm [16]:

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^d w_j \\ \text{subject to} \quad & w \cdot f_k \geq 1, \quad k = 1, \dots, m \quad (1) \\ & w_j \geq 0, \quad j = 1, \dots, d. \quad (2) \end{aligned}$$

B. Meta-LMA

This is an algorithm inspired by the LMA algorithm sharing the same goal ($y(v_k) > y(u_k)$) and the principal assumption that the subjective entertainment preference function (y) is a linear combination of individual features. According to meta-LMA, the weight vector $w = (w_1, w_2, \dots, w_d)$ of d selected features is adjusted to solve the following linear programming problem:

$$\begin{aligned} \text{maximize} \quad & \frac{1}{m} \sum_{k=1}^m g(y(f_k), \epsilon) \quad (3) \\ \text{subject to} \quad & y(f_i) \geq \delta, \quad i = 1, \dots, m \quad (4) \end{aligned}$$

where $y(f_k) = y(v_k) - y(u_k)$, δ is 0.05 in all experiments presented here, $g(y(f_k), \epsilon) = 1/(1 + e^{-\epsilon y(f_k)})$ is the sigmoid function, and $\epsilon = 30$ if $y(f_k) > 0$ and $\epsilon = 5$ if $y(f_k) < 0$. Both the sigmoidal shape of the objective function and its selected ϵ values are inspired by its successful application as a fitness function in neuroevolution preference learning problems on Playware test-bed games [17]. Thus, meta-LMA can be seen as a LMA variant enhanced with a nonlinear sigmoid objective function under maximization.

C. BL

As in LMA and meta-LMA, the main idea of the Bayesian-learning (BL) approach is that there is an unobservable function $y(x)$ associated with each instance (subject) x and that these function values express the preference relations observed in the data set. We utilize the approach of Williams and Rasmussen [18], who assume that the function values are random variables in a zero-mean Gaussian process. The process can be estimated using the covariance matrix. The covariance between the y -function values that correspond to instances x_i and x_j can be defined by the Gaussian kernel, and the elements of the $n \times n$ covariance matrix \mathcal{G} are given by $\mathcal{G} = K(x_i, x_j) =$

$\exp(-\kappa/2 \sum_{l=1}^d (x_i^l - x_j^l)^2)$, where κ is a positive constant and x_i^l denotes the l th individual feature value of subject x_i . Regarding the likelihood function, we used the proposal of Chu and Ghahramani [4] in order to portray the preference relations. The function is defined as follows:

$$P(\mathcal{L}|\mathcal{G}) = \prod_{k=1}^m P(v_k \succ u_k | y(v_k), y(u_k)).$$

Based on Bayes' theorem, the posterior probability can be written as

$$P(\mathcal{G}|\mathcal{L}) = \frac{P(\mathcal{G})}{P(\mathcal{L})} \prod_{k=1}^m P(v_k \succ u_k | y(v_k), y(u_k)).$$

Upon calculation of $P(\mathcal{G}|\mathcal{L})$, a gradient-based optimization algorithm is implemented to shape the Bayesian model. See [4] for further details of the BL algorithm.

D. Evolving ANNs

Given the high level of subjectivity of human preferences and the noisy nature of input data, we believe that more complex nonlinear functions such as those represented by ANNs might serve our purposes better. Thus, feedforward multilayered neural networks for learning the relation between the selected player features (ANN inputs) and the "entertainment value" (ANN output) of a game are also evaluated here. Since there are no prescribed target outputs for the learning problem (i.e., no differentiable output error function), ANN training algorithms such as backpropagation are inapplicable. Learning is achieved through artificial evolution.

The sigmoid function is employed at each neuron, and the connection weights take values from -5 to 5 to match with input values that are normalized into $[0, 1]$. In an attempt to minimize the ANN size, it was determined that a single hidden-layered architecture, containing 20 hidden neurons, is capable of successfully obtaining solutions of high fitness. This was determined by considering the performance of ANN architectures with up to two hidden layers containing up to 30 hidden neurons each.

1) *Genetic Algorithm*: A generational genetic algorithm (GA) [19] is implemented, which uses a fitness function that measures the difference between the children's reported preferences and the relative magnitude of the corresponding model output values y 's. The ANN is itself evolved. In the algorithm presented here, the ANN topology is fixed, and the GA chromosome is a vector of ANN connection weights. The algorithm is described only briefly here since it was previously presented in full detail in [6].

A population of N (1000 in this paper) networks is initialized randomly. Initial real values that lie within $[-5, 5]$ are picked for their connection weights randomly from a uniform distribution. Then, at each generation, the following are implemented. First, each member (neural network) of the population is given two d -tuples (where d is the number of game or player features), i.e., one for gameplay instance A and one for instance B for

each pair j of games compared by the player—see Section V and [6] for how these preferences are elicited from the players. In each case, the network returns two output values, namely, $y_{j,A}$ and $y_{j,B}$, representing the level of “fun” in each game. Second, each member of the population is evaluated via a fitness function that measures the degree of match between ANN outputs (y 's) and children's reported answers (see [6]). A high fitness results if the ranking of $y_{j,A}$ and $y_{j,B}$ matches the expressed preference of the children for each game pair j . Third, a fitness-proportional selection method is used. Last, Montana and Davis [20] crossover and Gaussian mutation are applied (see [6]).

The algorithm is terminated when either a good solution is found ($f > 0.95f_{\max}$, where f_{\max} is the maximum possible fitness) or a large number of generations (10 000) have been completed.

IV. TEST-BED PHYSICAL GAMES

The Playware [21] prototype playground consists of several building blocks (i.e., tangible tiles) that allow for the game designer (e.g., the child) to develop a significant number of different games within the same platform. The overall technological concept of Playware is based on physically implemented computational agents (the tiles) incorporating processing power, communication, input, and output. The *DigiWall* [22] and *Age Invaders* [23] mixed-reality (i.e., real environment augmented with virtual components) systems, the *Scorpiodome* [24] game system, the *STARS* [25] tabletop game, and the *PingPongPlus* [26] digitally enhanced ping-pong game are platforms closely related to Playware. See [6], [8], and [21] for further details on the Playware playground.

The “Bug-Smasher” game is used as the test-bed game in the experiments presented here. Bug-Smasher is developed on a 6×6 square-tile topology. During the game, different “bugs” (colored lights) appear on the game surface and disappear sequentially after a short period by turning a tile's light on and off, respectively. A bug's position is picked randomly according to a predefined level of the bugs' spatial diversity. The child's goal is to smash as many bugs as possible by stepping on the lighted tiles. Bug-Smasher has been used as a test bed in previous work; further details can be found in [6], [8], [17], and [27].

As previously mentioned, the overall goal of this paper is to model the reported entertainment preferences of users of augmented-reality physical games, such as Playware games. Modeling of entertainment preferences is grounded on features extracted from data recorded during play, which include the time when a tile was pressed, the pressure force on this tile, the state of the pressed tile (light's color), the speed of the game, and the entropy of bugs' visits. Four different preference-learning approaches are applied and compared. These approaches are used to learn the mapping between the selected features of the game and individual gameplay and the reported entertainment preferences. The generated model predicts the entertainment value of the game in real time, given the current state of the game and specific personal gameplay features of the user.

V. EXPERIMENTAL SETUP

According to the experimental design proposed in [8] and [28] for effectively capturing the level of entertainment, the test-bed game under investigation is played in variants. For this purpose, different states (e.g., “low” and “high”) of quantitative estimators of qualitative entertainment factors (e.g., challenge, curiosity, and fantasy [29]) are used. (The reader may refer to [8] and [30] for an analysis of quantitative measures of the challenge and curiosity factors for the Bug-Smasher game.) The combination of states/number of entertainment factors generates a pool of dissimilar games for the designer to investigate.

By experimental design (see [5] and [6]), each subject plays against k of the selected n variants of the chosen game in all permutations of pairs. (k equals two and n equals nine in the main experiment presented in this paper.) Thus, C_k^n is the required number of subjects to cover all combinations of k out of n game variants. More specifically, each child plays games in pairs (games A and B)—differing in the levels/states of one or more of the selected entertainment factors—for a selected time window. Each time a pair of games (“game pair”) is finished, the child is asked whether the first game was more “fun” than the second game (pairwise preference) or vice versa. Children are not interviewed but are asked to fill in a questionnaire, minimizing the interviewing effects reported in [31].

To minimize any potential order effects, each child is required to play the aforementioned games in the inverse order too. Statistical analysis of the effect of the order of gameplay on children's judgement of entertainment tests for both order effects and the level of inconsistency in children's preferences (see Section V-A1). For example, for a given pair (A, B), expressed preferences including both $A \succ B$ and $B \succ A$ may indicate a consistent preference for the first game of a pair (order effect) or an inability to rank the two games because they generate the same level of entertainment.

All subjects are given the same instructions by an experimenter who is unaware of the purpose of the experiment. No further oral or eye-contact communication takes place during experiment tasks and questionnaire, minimizing experimenter-expectancy effects [32]. The playing time window chosen (90 s in this paper) is a compromise between effective data collection (long-enough subject–game interaction to support a relative judgement) and not overstretching children with excessive periods of energetic physical play.

Capture of emotional state, such as entertainment, is generally considered a hard problem mainly because understanding emotion is hard [33]. Capturing reports of playing experiences or emotions is still tough since data obtained embed experimental noise and subjectivity. As previously mentioned, a pairwise preference scheme [2-alternative forced choice (2-AFC)] is used in self-reports of children. 2-AFC offers several advantages for a subjective entertainment capture: It minimizes the assumptions made about subjects' notions of “fun” and allows a fair comparison between the answers of different people. Since the focus is to construct a model relating reported entertainment preferences to individual playing features—which generalizes over the reports of different players—2-AFC is preferred to a ranking approach [31].

A. Main Experiment

Seventy-two normal-weighted (based on their body mass index) children whose ages cover a range between eight and ten years participated in the main experiment presented here. In this experiment, nine Bug-Smasher game variants were designed as a combination of two entertainment factors: challenge (bugs' speed) and curiosity (bugs' spatial diversity) of three states each ("low," "average," and "high"). The 72 children were asked to play a pair of Bug-Smasher variants according to the protocol presented in Section V. The Bug-Smasher test-bed game has already been used in previous work [6], [27]; however, the results presented in this paper are derived from new and more extensive experiments on this test bed.

Out of the total number of 288 games played in this experiment, in 274 games (137 game pairs), child–game interaction data were properly recorded. In the remaining 14 games, data were lost because of hardware failure (data transmission) during the game. There is no reason to suppose that the hardware failure, resulting in loss of data in 4.8% of the experimental games, has any particular bias with respect to our experimental hypothesis. The set of correctly recorded data collected from 137 game pairs of the main experiment underlies the analysis presented in this paper.

1) *Order Effects*: To check whether the order of playing Playware games affects the children's judgement of entertainment, we follow the order testing procedure introduced in [8], which is based on the times that the subject prefers the first or the second game in both pairs. This statistical analysis, which is presented in [17], shows that no significant order effect is detected: Statistical correlation of order of play with respect to reported entertainment (r_c) equals 0.102; the binomial distributed probability of this correlation to occur at random (p -value) equals 0.224. The reported insignificant order effect also demonstrates in part that effects such as a child's possible preference for the very first game played and the interplay between reported entertainment and familiarity with the game are statistically insignificant.

B. Controlled Physical Activity Experiment

In order to discover whether there is anything in the type of physical activity that is characteristic of an entertaining game, we designed an additional experiment where the physical-activity control is achieved through a nonentertaining variant of the Bug-Smasher game named the "Stomping game." This game requires children to stamp on Playware tiles in a similar way to Bug-Smasher but systematically and without any variation in response from the game platform. Children almost unanimously perceive this as boring. Experiments on this game were first introduced in [8].

For the control experimental protocol, we asked 18 naive normal-weighted children (nine boys and nine girls) aged eight to ten years to play five games each on the Playware platform. The set of five games played comprised five games of Bug-Smasher, in two pairs as in the experiment described in Section V, and a physical activity control game. As in the main experiment, two game variants of different levels of entertainment features (challenge and curiosity) were played in both

orders, giving four Bug-Smasher variant games plus the control game. All details regarding the protocol of the experiment follow the principles of the experimental setup described in Section V. Moreover, the statistical analysis presented in [8] demonstrates that the order of play does not significantly affect children's preferences ($r_c = -0.166$ and p -value = 0.2025) expressed in the experiment. The reader may refer to [8] for further details of the controlled physical activity experiment since its detailed presentation does not lie within the main scope of this paper.

All four preference learning approaches presented in Section III are trained and validated on data obtained from the main experiment (see Section V-A), namely, "main data set." The best approach is then evaluated using unseen data from the physical activity control experiment to determine the extent to which the constructed user model generalizes. Data from this experiment are referred to as "control data set" in the following sections.

VI. EXTRACTED FEATURES

Since, with the current implementation of the Playware platform, the only input to the system is through a force-sensing-resistor (FSR) sensor, quantitative individual playing characteristics can only be based on three measurable features: the state (position and LED color) of a pressed tile, the time that a pressed-tile event took place, and the pressure force on a pressed tile. Pressed-tile events are recorded in real time, and a selection of nine player interaction features are calculated for each child. These include the number of smashed bugs over the total number of bugs appeared P (i.e., child's score), the number of interactions with the game environment N_I , the average and the variance of the response times ($E\{r_t\}, \sigma^2\{r_t\}$), the average and the variance of the distance between the pressed tile and the bugs appearing on the game ($E\{D_b\}, \sigma^2\{D_b\}$), the average and the variance of the pressure recorded from the FSR sensor ($E\{p\}, \sigma^2\{p\}$), and the entropy of the tiles that the child visited H_C . The complete set of candidate features also includes the quantitative controllable game features of challenge (S) and curiosity (H).

A. Main Data-Set Statistical Analysis

The analysis presented in [17] identifies statistically significant correlations between children's notion of entertainment and a few of the aforementioned individual player features and/or the quantitative entertainment factors (game features): challenge and curiosity. The obtained significant effects of $E\{r_t\}$, N_I , and $E\{p\}$ appear to be commonsensical since the Bug-Smasher game belongs to the genre of action physical games where the level of engagement of the user tends to have a significant effect on the number of interactions and the reaction time of the player [34]. In Bug-Smasher, the more a child is entertained, the more (N_I) and harder ($E\{p\}$) she/he tends to interact with the game platform. This behavior generates lower average response time ($E\{r_t\}$) and higher average pressure on the tiles ($E\{p\}$). The obtained highly significant effect of $E\{r_t\}$ is consistent with previous experiments on the Bug-Smasher game [6].

On the other hand, it appears that reported entertainment does not straightforwardly correlate with the levels of challenge ($r_c = 0.0$, p -value = 0.5382) and curiosity ($r_c = 0.0909$ and p -value = 0.1448). This suggests that the relation between those two factors and entertainment is nonlinear, i.e., too much or too little of each is not entertaining. This effect motivates the need of a level of personalization (measured by the other interaction features) included as a factor in entertainment modeling. The feature selection procedure presented in Section VII helps the designer choose specific individual player features that can successfully generate a model that maps between children's behavior, game features, and reported entertainment.

B. Control-Data-Set Statistical Analysis

By following the statistical analysis procedure presented in [17] for the control data set, we obtain significant effects of H , H_C , and $\sigma^2\{D_b\}$ to reported entertainment. It therefore appears that the higher the spatial diversity of the bugs H and children H_C and the higher the variance of the distances between the child and the bugs ($\sigma^2\{D_b\}$), the higher the reported entertainment. These effects that differ from those derived from the main experiment data are nevertheless expected since the Stomping game is now included in our investigation. The features mentioned are those whose values are, in general, much higher in the preferred Bug-Smasher game variants than in the generally nonpreferred Stomping game.

VII. PREFERENCE LEARNING WITH FEATURE SELECTION

Two different input feature selection schemes are used to pick the appropriate feature subset that generates the highest classification performance between preferred and nonpreferred games. Given the set of candidate features described in Section VI, the nBest and SFS methods are applied. The nBest method picks the n individually best features (with regard to a performance function) from the feature subset. The SFS method, by contrast, is a bottom-up search procedure where one feature is added at a time to the current feature set. The feature to be added is selected from the subset of the remaining features so that the new feature set generates the maximum value of the performance function over all candidate features for addition [35].

The SFS method is tested since it has been successfully applied in a wide variety of feature selection problems yielding high-performance values with minimal feature subsets (see, for example, [36] for further discussion and application to the classification problem of process identification in resistance spot welding). On the other hand, the nBest method is included for comparative purposes, with it being the most popular technique for feature selection. More advanced methods, such as sequential floating forward search and Fisher projection (FP) [33] could be used in future experiments, and results could be compared to the existing studies. The feature selection procedure followed here evaluates the usability of each one of the features available and obtains the minimal feature subset approximation to the feature subset that performs best in the classification between preferred and nonpreferred games.

TABLE I
VALIDATION PERFORMANCE P_v (IN PERCENT) OF THE INDIVIDUAL PLAYER AND GAME FEATURES OF THE FOUR PREFERENCE LEARNING MECHANISMS. $E\{r_t\}$ AND $\sigma^2\{r_t\}$ ARE THE AVERAGE AND THE VARIANCE OF THE RESPONSE TIME, RESPECTIVELY; $\sigma^2\{D_b\}$ IS THE VARIANCE OF THE DISTANCES BETWEEN THE PRESSED TILE AND THE BUGS APPEARING ON THE GAME; N_I IS THE TOTAL NUMBER OF INTERACTIONS; H IS THE QUANTITATIVE MEANS FOR THE CONTROLLABLE GAME FEATURE OF CURIOSITY; $E\{p\}$ IS THE AVERAGE PRESSURE FORCE RECORDED FROM THE FSR SENSOR; H_C IS THE ENTROPY OF THE TILES THAT THE CHILD VISITED; $\sigma^2\{p\}$ IS THE VARIANCE OF THE PRESSURE FORCES RECORDED FROM THE FSR SENSOR; $E\{D_b\}$ IS THE AVERAGE DISTANCE BETWEEN THE PRESSED TILE AND THE BUGS APPEARING ON THE GAME; S IS THE QUANTITATIVE MEANS FOR THE CONTROLLABLE GAME FEATURE OF CHALLENGE; AND P IS THE SCORE

	LMA		meta-LMA		BL		Evolving ANN	
Feature	P_v	Feature	P_v	Feature	P_v	Feature	P_v	
$E\{r_t\}$	60.00	$E\{r_t\}$	60.00	N_I	69.50	$E\{r_t\}$	62.22	
S	56.67	$\sigma^2\{p\}$	57.78	$E\{r_t\}$	68.24	$\sigma^2\{r_t\}$	61.11	
$\sigma^2\{r_t\}$	54.44	N_I	55.56	$E\{p\}$	63.19	$\sigma^2\{D_b\}$	56.67	
$E\{D_b\}$	54.44	$\sigma^2\{r_t\}$	54.44	$\sigma^2\{r_t\}$	60.65	N_I	56.67	
$\sigma^2\{D_b\}$	51.11	H	54.44	H	55.60	H	55.56	
P	50.00	$E\{p\}$	54.44	$\sigma^2\{p\}$	50.55	$E\{p\}$	53.33	
H_C	48.89	$E\{D_b\}$	54.44	H_C	39.17	H_C	53.33	
H	45.56	H_C	51.11	$E\{D_b\}$	14.42	$\sigma^2\{p\}$	51.11	
$E\{p\}$	45.56	$\sigma^2\{D_b\}$	51.11	P	12.63	$E\{D_b\}$	51.11	
N_I	44.44	P	50.00	S	11.37	S	46.67	
$\sigma^2\{p\}$	42.22	S	38.89	$\sigma^2\{D_b\}$	10.65	P	43.33	

Feature selection algorithms determine the features under investigation for each learning mechanism; in the case of the evolving ANN, the selected features define its input vector. To evaluate the performance of each feature subset considered by each selection algorithm, the available data are randomly divided into training and validation data sets consisting of two-thirds and one-third of the data, respectively. The performance of each user model is measured through the average classification accuracy of the model in three independent runs using the threefold cross-validation technique on the training and validation data sets. Since we are interested in the minimal feature subset that yields the highest performance, we terminate the feature selection procedure (nBest or SFS) when an added feature yields equal or lower validation performance than the performance obtained without it. Feature selection applied this way resembles a hill-climber approach that is at risk of being trapped by local optimality. Furthermore, since cross-validation is used for assessing the performance of each feature subset, the validation data set is not used here for early stopping of the training procedure to limit overfitting.

A. Best Feature Selection

The experiment presented here tests the validation performance of single individual player and game features. Given the selected feature set, all four approaches presented in Section III-D1 are followed and evaluated (see Section VII). The training and validation performances of each of the individual player and game features for each learning mechanism are presented in Table I, where features are ranked by validation performance.

TABLE II
nBEST FEATURE SELECTION METHOD: CLASSIFICATION ACCURACY
(IN PERCENT) OF LMA, META-LMA, BL, AND ANN APPROACHES.
THE HIGHEST PERFORMANCE OF EACH APPROACH OBTAINED
THROUGH A MINIMAL FEATURE SUBSET \mathcal{F} APPEARS IN BOLD

LMA		Meta-LMA		BL		ANN	
\mathcal{F}	P_v	\mathcal{F}	P_v	\mathcal{F}	P_v	\mathcal{F}	P_v
$E\{r_t\}$	60.00	$E\{r_t\}$	60.00	N_I	69.50	$E\{r_t\}$	62.22
S	57.41	$\sigma^2\{p\}$	62.22	$E\{r_t\}$	70.21	$\sigma^2\{r_t\}$	58.88
$\sigma^2\{r_t\}$	58.02	N_I	63.33	$E\{p\}$	61.92	$\sigma^2\{D_b\}$	44.44
$E\{D_b\}$	60.00	$\sigma^2\{r_t\}$	62.22	$\sigma^2\{p\}$	48.95	N_I	46.67

TABLE III
SFS FEATURE SELECTION METHOD: CLASSIFICATION ACCURACY
(IN PERCENT) OF LMA, META-LMA, BL, AND ANN APPROACHES.
THE HIGHEST PERFORMANCE OF EACH APPROACH OBTAINED
THROUGH A MINIMAL FEATURE SUBSET \mathcal{F} APPEARS IN BOLD

LMA		Meta-LMA		BL		ANN	
\mathcal{F}	P_v	\mathcal{F}	P_v	\mathcal{F}	P_v	\mathcal{F}	P_v
$E\{r_t\}$	60.00	$E\{r_t\}$	60.00	N_I	69.50	$E\{r_t\}$	62.22
$\sigma^2\{r_t\}$	63.33	P	67.78	$\sigma^2\{D_b\}$	73.30	$\sigma^2\{p\}$	67.77
$\sigma^2\{D_b\}$	62.22	$\sigma^2\{r_t\}$	70.00	$\sigma^2\{r_t\}$	64.45	H	68.88
$E\{p\}$	62.22	H_C	70.00	P	55.38	N_I	82.22

The impact of the recorded response times (r_t) is demonstrated in Table I. More specifically, the average of these values generates the highest cross-validation performances when LMA, meta-LMA, and evolving ANN are applied (see also [6] for the impact of $E\{r_t\}$ on reported entertainment in the Bug-Smasher game). On the other hand, BL reports the number of interactions with the playground as the single feature that best predicts reported entertainment. The high cross-validation accuracy of N_I (69.50%) suggests that the BL mechanism generates a nonlinear mapping of N_I to reported entertainment—better than that of $E\{r_t\}$ —which was discovered by neither the linear (LMA and meta-LMA) nor the nonlinear (evolving ANN) IPL approaches. Overall, results obtained show the inability of any single feature to predict reported entertainment in Bug-Smasher. Given that the best performed feature (N_I) yields a cross-validation performance of 69.50%, it is apparent that more features are required to effectively model children’s notion of entertainment.

B. More Features

The initial feature subset for all four preference learning approaches includes the feature that performs best in the single-feature experiment ($E\{r_t\}$ for LMA, meta-LMA, and evolving ANN; N_I for BL). By applying the nBest and SFS methods for each learning approach, we obtain cross-validation performances presented in Tables II and III, respectively. Comparing feature selection methods within each preference learning approach, the SFS method (see Table III) generates feature subsets that yield higher validation performance than feature subsets generated by nBest (see Table II) for all four preference-learning mechanisms. This confirms the benefits of searching for appropriate feature sets with SFS.

As expected, results obtained through SFS show the advantage of nonlinear (BL and ANN) over linear (LMA and meta-LMA) learning approaches for our preference learning case study. Even though the LMA method compared to neuroevolution has generated higher performing solutions in specific

problems (e.g., routing [3]), the inverse occurs in our paper. Within nonlinear approaches, the comparative study between BL and evolving ANN is introduced in this paper and shows that the latter outperforms the former for our problem. More specifically, the best cross-validation performance (82.22%; average of 76.66%, 80.00%, and 90.00%) is achieved when the ANN input contains $E\{r_t\}$, $\sigma^2\{p\}$, H , and N_I , while adding more features to the subset does not yield significantly higher performance. On the other hand, the best performances that LMA, meta-LMA, and BL achieve are significantly lower, i.e., 63.33%, 70.00%, and 73.30%, respectively. However, it is worth noticing that the highest performance of BL (73.30%) is achieved by considering two (N_I , $\sigma^2\{D_b\}$) instead of four features (evolving ANN). This indicates the promise of BL as an IPL mechanism for the specific case study. These mechanisms’ generated performance can be used as a baseline for comparison to the best evolved ANN solution, which indicates that the nonlinear combination of $E\{r_t\}$, $\sigma^2\{p\}$, H , and N_I is sufficient for an efficient predictor of reported entertainment for the tested Playware games.

The best performance obtained (82.22%) appears to be rather low. However, the binomial-distributed probability of this performance to occur at random (**0.0003**) is statistically significant and indicates that ANNs are capable of successfully predicting children’s reported entertainment preferences based on features extracted from individual gameplay data, despite the difficulty of the problem. (For comparison, the average classification performance of 30 ANNs with random weights is 49.77%.)

Difficulties in obtaining higher classification accuracy are caused in part by experimental noise in the recorded features and in part by inconsistency (due to the 2-AFC protocol) in children’s answers in self-reports. Even though comparative fun analysis is a reliable and established method for capturing reported entertainment in computer [37] and augmented-reality [6] games, it generates a significant amount of uncertainty in subjects’ reported answers. Uncertainty appears when the two games played are not significantly different with regard to the entertainment value that they generate for the player and therefore cannot be easily distinguished.

1) *Evolved ANN*— $\{E\{r_t\}, \sigma^2\{p\}, H, N_I\}$ *Feature Subset*: A more detailed analysis of the evolved ANN model that yields the best classification accuracy is presented here. Fig. 1 shows the highest performing trained ANN output with regard to $\sigma^2\{p\}$ and N_I for six points in the ($E\{r_t\}$, H) search space. These values constitute the combinations of two $E\{r_t\}$ states (0 and 1 named *fast* and *slow*, respectively) and the three states used for H (0.33, 0.66, and 1 named *low*, *average*, and *high*, respectively). The aforementioned presentation helps toward interpreting the mapping between $\sigma^2\{p\}$, N_I , and reported entertainment according to how fast children react with the playground and the level of curiosity. As in the results presented in [7], [8], and [17], all three fittest ANNs generated, each trained on different sets comprising two-thirds of the total data, exhibit similar qualitative features to the surfaces shown in Fig. 1.

As seen from Fig. 1, fast children ($E\{r_t\} = 0$) appear to enjoy average and high curiosity values, except when high N_I

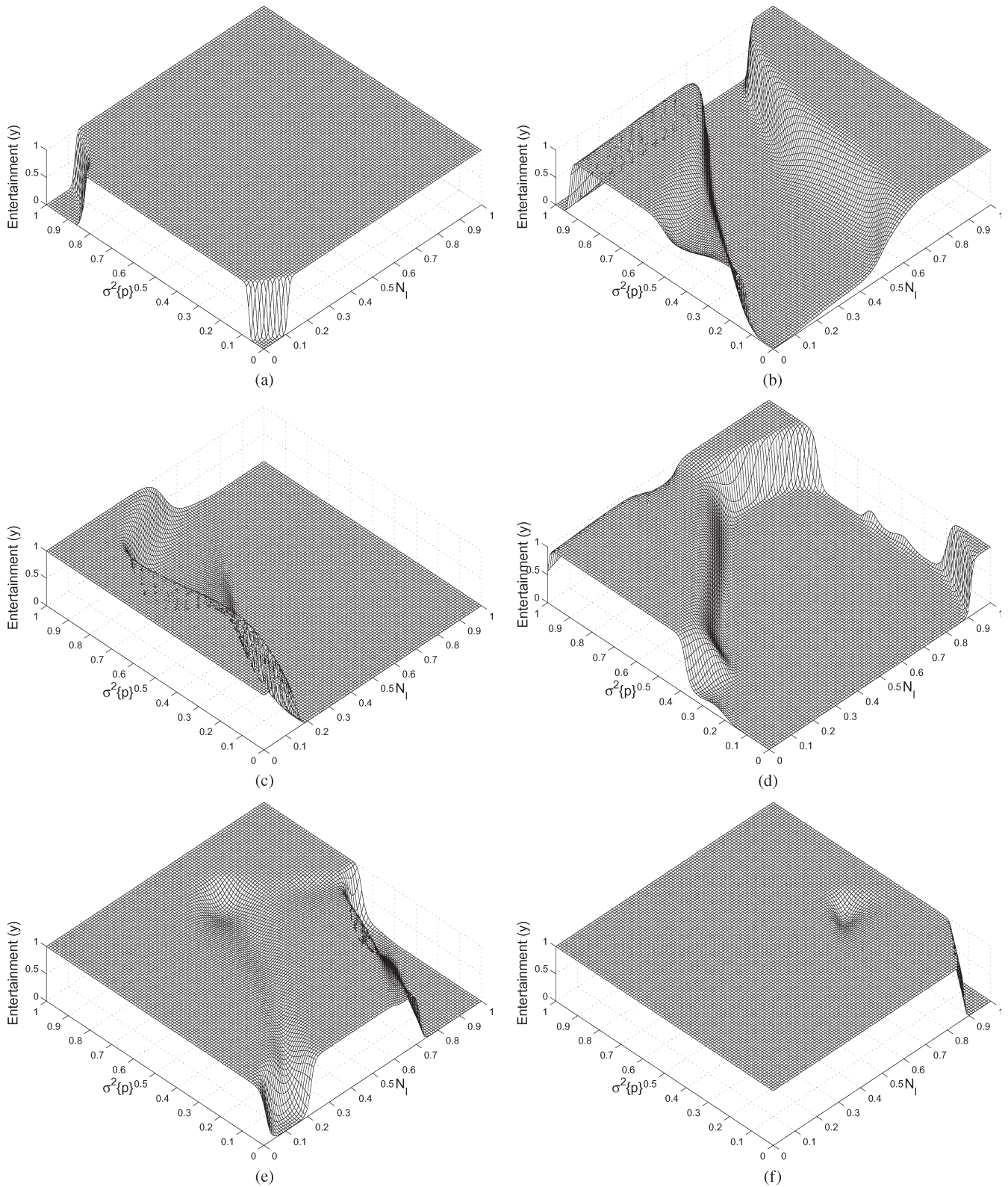


Fig. 1. Trained ANN ($f = 95.85$) that yields the highest validation performance (90.00%): ANN output y (entertainment value) with regard to $\sigma^2\{p\}$ and N_I for all six combinations of two $E\{r_t\}$ states (slow and fast) and three H states (low, average, and high). (a) $E\{r_t\} = 1.0$ (slow) and $H = 0.33$ (low). (b) $E\{r_t\} = 1.0$ (slow) and $H = 0.66$ (average). (c) $E\{r_t\} = 1.0$ (slow) and $H = 1.0$ (high). (d) $E\{r_t\} = 0.0$ (fast) and $H = 0.33$ (low). (e) $E\{r_t\} = 0.0$ (fast) and $H = 0.66$ (average). (f) $E\{r_t\} = 0.0$ (fast) and $H = 1.0$ (high).

values ($N_I > 0.7$) are combined with low values of $\sigma^2\{p\}$ [$\sigma^2\{p\} < 0.4$ —see Fig. 1(e) and (f)]. Fast children’s preference for low levels of curiosity is met only when their behavior gen-

erates high values (approximately greater than 0.5) of $\sigma^2\{p\}$, independently of the N_I value [see Fig. 1(d)]. On the other hand, slow children appear to prefer low curiosity levels, except

TABLE IV
EVOLVED ANNs (FEATURE SUBSET: $\{E\{r_t\}, \sigma^2\{p\}, H, N_I\}$) TRAINED ON THE MAIN DATA SET: CLASSIFICATION ACCURACY (IN PERCENT) ON UNSEEN CONTROL DATA SET. THE AVERAGE CLASSIFICATION PERFORMANCE OF TEN NETWORKS WITH RANDOM WEIGHTS IS GIVEN FOR COMPARISON (RANDOM NETWORK)

	Bug-Smasher game vs. Stomping game	Preferred game vs. non-preferred game	Total
Random network	54.34	52.43	53.39
$\{E\{r_t\}, \sigma^2\{p\}, H, N_I\}$	73.91	59.49	66.66

when the N_I value that they generate is low ($N_I < 0.1$) and combined with either very high (> 0.9) or very low (< 0.1) $\sigma^2\{p\}$ values [see Fig. 1(a)]. Average curiosity levels are preferred by slow children in many fewer cases, i.e., when their N_I value is low and the $\sigma^2\{p\}$ value is high or when their N_I value is high and the $\sigma^2\{p\}$ value is low [see Fig. 1(b)]. Finally, high curiosity is rarely preferred by slow children, and this occurs only when their N_I values are low, independently of their $\sigma^2\{p\}$ value [Fig. 1(c)].

The obtained effects of curiosity in reported entertainment are consistent in part with previous studies on the Bug-Smasher game [6]. In that study, the relation between challenge, curiosity, and average response time was reported through a lower scale experiment of 28 children. It was found that fast children liked games, independently of curiosity, whereas children reacting slowly to game events preferred games of low curiosity levels.

2) *Validation on Control Data Set:* To investigate the extent to which the predictive model of entertainment preference computed using data from the main experiment generalizes to new experimental data, the best performing evolved ANNs presented in Section VII-B1 are presented with and evaluated on the unseen game and data of the physical activity control experiment.

Table IV shows the average total classification accuracy (fourth column) and the subclassification performance for the comparisons between the Bug-Smasher game played and the Stomping game (second column), as well between the Bug-Smasher game chosen as more entertaining and the Bug-Smasher game chosen as less entertaining (third column) of all three evolved ANNs.

The performance obtained equals 66.66%, which appears rather low compared to 82.22% of correct matching on the validation data of the main experiment. However, the reported complexity of the task [17] and the binomial-distributed probability of this performance to occur at random (**0.0136**) indicate that the ANN is a quite effective, robust, and generic predictor of children's reported entertainment preferences based on their interaction data with the physical-game platform. The average performance of ten ANNs identical in structure to the evolved ones, but with random weights, is given for comparison.

VIII. CONCLUSION AND DISCUSSION

This paper explored the interplay between individual playing behaviors and children's entertainment preferences in physical play. More specifically, the quantitative impact of children's entertainment (known from self-reports) on children's gameplay statistics was investigated through an action game (Bug-

Smasher) developed on the Playware playground. Features extracted from the interaction between the user (child) and the physical-game platform may provide a means for distinguishing between entertaining and nonentertaining games, as well as between gaming activities and gamelike physical activities (stomping).

Four IPL mechanisms—both linear (LMA and BL) and nonlinear (meta-LMA and neuroevolution)—were applied to the problem of predicting entertainment preference using gameplay features. Two feature selection methods (nBest and SFS) were used to determine which subset of features should be used as model inputs. Of the two, SFS was found to generate better performance than nBest for all learning mechanisms. Neuroevolution produced the best performance of the four IPL mechanisms.

SFS selection of individual and game features derived from the data of the main experiment extracted a feature subset including the average response times ($\{E\{r_t\}$), the variance of pressure forces ($\sigma^2\{p\}$), the number of interactions (N_I), and the controllable game feature of curiosity H . These inputs feed an ANN model that correctly predicts the reported entertainment preferences of children with a cross-validation accuracy of 82.22% on the unseen data of the main experiment.

Moreover, a cross-validation performance of 66.66% is obtained when evaluating the ANN trained on data from the main experiment using unseen data from the control experiment. Even though the obtained performance appears low, its binomial-distributed probability to occur at random is **0.0136**. One has to consider the difficulty of classifying accurately reported emotions through gameplay data [17] and the fact that the model is trained on data from a group of children whose individual preferences may not be mutually consistent.

The results presented indicate that the evolved ANNs are nevertheless successful predictors of children's reported entertainment grounded on their gameplay data and validate the hypothesis that there are personalized gameplay features that capture entertainment in physical games.

As far as the experiment protocol is concerned, a 4-AFC approach could be adopted for future protocol design. Children will then be able to choose among the following alternatives: One game is more "fun" than the other (2-AFC), both games are equally "fun," or neither game was "fun." This protocol provides similar information for the machine learning process while eliminating the inconsistency generated by 2-AFC when two game variants are indistinguishable.

We believe that the entertainment modeling approach through the individual playing and game features presented here is general over the majority of action games that could be created with Playware. Moreover, it is our belief that the

entertainment models proposed here may very well be applied to other interactive entertainment systems that include physical activity. However, each game demonstrates individual entertainment features that may have an impact on the child's gameplay, and therefore, new games need to be tested to confirm the generality of the approach.

Individual differences in children's preferences and playing behavior generate difficulties in generalization over subjects. This is a fundamental limitation of attempting to construct a model based on combined data from multiple subjects: a game that is exciting and fun to one child may be too fast or too slow to another. Nevertheless, the results presented show that some generalization across individuals is possible, in that the evolved ANNs do predict children's preferences with reasonable performance. With further work, it may be possible to improve performance, for example, by clustering individual players into classes, depending on the observed playing style. Each class could then have its own model, simplifying the machine learning problem.

The advantages of neuroevolution over other preference-learning techniques presented here appear to be primarily due to the nature of data obtained. Data are characterized by experimental noise that is apparent in the recorded features (ANN input) and inconsistency in children's answers in self-reports. Given the ability of ANNs to generalize over noisy training data, we believe that neuroevolution might serve well for constructing efficient cognitive and affective models built on data that embed significant levels of noise (e.g., physiological signals). Studies in the aforementioned fields have already demonstrated the benefits of the use of neuroevolution for yielding successful user models [8], [38].

The proposed approach can be used for adaptation of the game's entertainment features (challenge and curiosity) according to the player's individual playing features in real time in physical games. The key to this is the observation that the models (e.g., ANNs) relate features to an entertainment value. It is therefore possible in principle to infer what changes to game features (challenge, curiosity) will cause an increase in the entertainment value of the game and to adjust game parameters to make those changes. For further discussion on this future direction, the reader may refer to [8] and [17].

REFERENCES

- [1] J. Fürnkranz and E. Hüllermeier, "Preference learning," *Künstliche Intelligenz*, vol. 19, no. 1, pp. 60–61, 2005.
- [2] J. Read, S. MacFarlane, and C. Cassey, "Endurability, engagement and expectations," in *Proc. Int. Conf. Interaction Des. Children*, 2002, pp. 1–23.
- [3] C.-N. Fiechter and S. Rogers, "Learning subjective functions with large margins," in *Proc. 17th ICML*, 2000, pp. 287–294.
- [4] W. Chu and Z. Ghahramani, "Preference learning with Gaussian processes," in *Proc. 22nd ICML*, 2005, pp. 137–144.
- [5] G. N. Yannakakis and J. Hallam, "Towards capturing and enhancing entertainment in computer games," in *Proc. 4th Hellenic Conf. Artif. Intell.*, vol. 3955, *Lecture Notes in Artificial Intelligence*, Heraklion, Greece, May 2006, pp. 432–442.
- [6] G. N. Yannakakis, H. H. Lund, and J. Hallam, "Modeling children's entertainment in the playware playground," in *Proc. IEEE Symp. Comput. Intell. Games*, May 2006, pp. 134–141.
- [7] G. N. Yannakakis and J. Hallam, "Preliminary studies for capturing entertainment through physiology in physical play," Maersk Institute, Univ. Southern Denmark, Copenhagen, Denmark, Tech. Rep. TR-2007-5.
- [8] G. N. Yannakakis, J. Hallam, and H. H. Lund, "Entertainment capture through heart rate activity in physical interactive playgrounds," *User Model. User-Adapted Interaction*, vol. 18, no. 1/2, pp. 207–243, Feb. 2008.
- [9] J. Doyle, "Prospects for preferences," *Comput. Intell.*, vol. 20, no. 2, pp. 111–136, May 2004.
- [10] R. Herbrich, T. Graepel, P. Bollmann-Sdorra, and K. Obermayer, "Learning preference relations for information retrieval," in *Proc. Text Categorization ICML*, 1998, pp. 80–84.
- [11] T. Joachims, "Optimizing search engines using clickthrough data," in *Proc. KDD*, 2002, pp. 133–142.
- [12] A. Bahamonde, G. F. Bayón, J. Díez, J. R. Quevedo, O. Luaces, J. J. del Coz, J. Alonso, and F. Goyache, "Feature subset selection for learning preferences: A case study," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, pp. 49–56.
- [13] X. Yao, "Evolving artificial neural networks," *Proc. IEEE*, vol. 87, no. 9, pp. 1423–1447, Sep. 1999.
- [14] M. T. Gervasio, M. D. Moffitt, M. E. Pollack, J. M. Taylor, and T. E. Uribe, "Active preference learning for personalized calendar scheduling assistance," in *Proc. 10th Int. Conf. IUI*, 2005, pp. 90–97.
- [15] M. Gervasio, W. Iba, and P. Langley, "Learning user evaluation functions for adaptive scheduling assistance," in *Proc. 16th Int. Conf. Mach. Learn.*, 1999, pp. 152–161.
- [16] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [17] G. N. Yannakakis and J. Hallam, "Game and player feature selection for entertainment capture," in *Proc. IEEE Symp. Comput. Intell. Games*, HI, Apr. 2007, pp. 244–251.
- [18] C. K. I. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Proc. Adv. Neural Inf. Process. Syst.*, 1996, pp. 598–604.
- [19] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: Univ. Michigan Press, 1975.
- [20] D. J. Montana and L. D. Davis, "Training feedforward neural networks using genetic algorithms," in *Proc. 11th IJCAI*, 1989, pp. 762–767.
- [21] H. H. Lund, T. Klitbo, and C. Jessen, "Playware technology for physically activating play," *Artif. Life Robot.*, vol. 9, no. 4, pp. 165–174, Dec. 2005.
- [22] M. Liljedahl and S. Lindberg, "DigiWall—An audio mostly game," in *Proc. Int. Conf. Auditory Display*, London, U.K., Jun. 20–23, 2006.
- [23] Mixed Reality Lab, Singapore. [Online]. Available: <http://www.mixedreality.nus.edu.sg/>
- [24] G. Metaxas, B. Metin, J. Schneider, G. Shapiro, W. Zhou, and P. Markopoulos, "SCORPIODROME: An exploration in mixed reality social gaming for children," in *Proc. ACM SIGCHI Int. Conf. Adv. Comput. Entertainment Technol.*, Jun. 2005, pp. 229–232.
- [25] C. Magerkurth, R. Stenzel, and T. Prante, "STARS—A ubiquitous computing platform for computer augmented tabletop games," in *Proc. Ext. Abstr. UbiComp*, 2003, pp. 267–268.
- [26] H. Ishii, C. Wisneski, J. Orbanes, B. Chun, and J. Paradiso, "PingPong-Plus: Design of an athletic-tangible interface for computer-supported cooperative play," in *Proc. SIGCHI Conf. Human Factors Comput. Syst. CHI*, 1999, pp. 394–401.
- [27] G. N. Yannakakis, J. Hallam, and H. H. Lund, "Capturing entertainment through heart-rate dynamics in the playware playground," in *Proc. 5th Int. Conf. Entertainment Comput.*, vol. 4161, *Lecture Notes in Computer Science*, Cambridge, U.K., 2006, pp. 314–317.
- [28] G. N. Yannakakis and J. Hallam, "Modeling and augmenting game entertainment through challenge and curiosity," *Int. J. Artif. Intell. Tools*, vol. 16, no. 6, pp. 981–999, Dec. 2007.
- [29] T. W. Malone, "What makes computer games fun?" *Byte*, vol. 6, pp. 258–277, 1981.
- [30] G. N. Yannakakis and J. Hallam, "Feature selection for capturing the experience of fun," in *Proc. AIIDE Workshop Optimizing Player Satisfaction*, G. N. Yannakakis and J. Hallam, Eds., 2007, pp. 37–42. AAAI Press Tech. Rep. WS-07-01.
- [31] R. L. Mandryk, K. M. Inkpen, and T. W. Calvert, "Using psychophysiological techniques to measure user experience with entertainment technologies," *Behav. Inf. Technol.*, vol. 25, no. 2, pp. 141–158, 2006.
- [32] R. Rosenthal, "Covert communication in laboratories, classrooms, and the truly real world," *Curr. Directions Psychol. Sci.*, vol. 12, no. 5, pp. 151–154, 2003.
- [33] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.
- [34] C. Beal, J. Beck, D. Westbrook, M. Atkin, and P. Cohen, "Intelligent modelling of the user in interactive entertainment," in *Proc. AAAI Spring Symp. Artif. Intell. Interactive Entertainment*, Stanford, CA, 2002, pp. 8–12.

- [35] P. Devijver and J. Kittler, *Pattern Recognition—A Statistical Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [36] E. Haapalainen, P. Laurinen, P. Junno, H. Tuovinen, and J. Roening, "Methods for classifying spot welding process: A comparative study of performance," in *Proc. IEA/AIE*, 2005, pp. 412–421.
- [37] G. N. Yannakakis, "AI in computer games: Generating interesting interactive opponents by the use of evolutionary computation," Ph.D. dissertation, Univ. Edinburgh, Edinburgh, U.K., Nov., 2005.
- [38] G. N. Yannakakis and J. Hallam, "Entertainment modeling through physiology in physical play," *Int. J. Human-Comput. Stud.*, vol. 66, no. 10, pp. 741–755, Oct. 2008.



Georgios N. Yannakakis (S'04–M'05) received the five-year Diploma in production engineering and management and the M.Sc. degree in financial engineering from the Technical University of Crete, Chania, Greece, in 1999 and 2001, respectively, and the Ph.D. degree in informatics from the University of Edinburgh, Edinburgh, U.K., in 2005.

He is currently an Associate Professor with the IT University of Copenhagen, Copenhagen, Denmark. Prior to joining the Center for Computer Games Research, IT University of Copenhagen, in 2007, he

was a Postdoctoral Researcher with the Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, Denmark. His research interests include artificial life, neuroevolution, emergent cooperation within multiagent systems, artificial intelligence in computer games, cognitive modeling, and affective computing. He has published around 40 journal and international conference papers on the aforementioned fields.



Manolis Maragoudakis (M'09) received the Diploma in computer science from the University of Crete, Chania, Greece, in 1999, and the Ph.D. degree in artificial intelligence from the University of Patras, Patras, Greece, in 2005.

He is currently an Assistant Professor with the Department of Information and Communication Systems Engineering, University of the Aegean, Karlovasi, Greece. Upon completion of his military service, he was a Postdoctoral Researcher with the University of Patras and a Lecturer with the Ionian

University, Corfu, Greece. His research interests include data mining, artificial intelligence, Bayesian methods, and user modeling. He has published around 45 journal and international conference papers on the aforementioned fields.



John Hallam received the B.A. degree in mathematics (with first-class honors) from the University of Oxford, Oxford, U.K., in 1979 and the Ph.D. degree from the University of Edinburgh, Edinburgh, U.K., in 1984.

He joined the Department of Artificial Intelligence, University of Edinburgh, as a Faculty Member in 1985. He established the Edinburgh Mobile Robotics Research Group, having been active in mobile robotics research for almost 20 years. In 2003, he moved to the Maersk Mc-Kinney Moller Institute,

University of Southern Denmark, Odense, Denmark. The current focus of his catholic research interest is in biological modeling using robotic techniques, evolutionary robotics, collective robotics, and modeling of user satisfaction in computer games. He has published around 100 journal and international conference papers on various robotic and nonsymbolic computing topics and has designed electronic hardware for research, teaching, and commercial purposes.

Dr. Hallam is the President of the International Society for Adaptive Behavior, a member of the London Mathematical Society, and a Director of 3 Lions Design Ltd., a small company that does commercial electronic design.