Computers in Human Behavior 49 (2015) 147–155

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Predicting expert–novice performance as serious games analytics with objective-oriented and navigational action sequences

Christian Sebastian Loh [a,*], Yanyan Sheng [b,1], I-Hung Li [c,2]

[a] Virtual Environment Lab (V-Lab), Southern Illinois University, Carbondale, IL, USA
[b] Quantitative Methods, Southern Illinois University, Carbondale, IL, USA
[c] Virtual Environment Lab (V-Lab), Southern Illinois University, Carbondale, IL, USA

A R T I C L E   I N F O

Article history:
Available online 14 March 2015

Keywords:
Similarity measures
Expert–novice performance
Serious games analytics
Action sequences
Partial least squares
Discriminant analysis

A B S T R A C T

Previous research differentiated expert vs. novice performances based on how (dis)similar novices' action sequences were from that of the expert's by way of similarity measures. Action sequences were coded using an 'objective-oriented' (or task-based) approach based on the sequence of objectives/tasks completed in-game. Findings from these studies suggest that the task-based similarity measures is a better predictor than (a) distance traversed, and (b) time (of completion).

In this study, we suggested an alternative method to code action sequences of experts and novices by way of a 'navigational' (or tile-based) approach. We divided a game-map into grids/tiles of different sizes to facilitate tracing of the path traversed by players in game and proceeded to test the effect of grid sizes on differentiating between experts and novices. We further compared the two different action sequence coding approaches and their abilities to measure players' competency improvement in serious games. The results of the study showed that the size of game grids does matter, and that both task-based and tile-based action sequence coding approaches are useful for serious games analytics.

1. Introduction

Generally speaking, the purpose of analytics is to discover value-added properties based on user-generated data. For instance, to stakeholders in the digital and mobile entertainment gaming industry, the purpose of game analytics is to create new revenue out of player-generated gameplay data (i.e., monetization). Since serious games was originally envisioned (Krulak, 1997) to be advanced training tools for the improvement of decision-making skills and job performance in trainees/learners, the purpose of serious games analytics would logically be to gain insights (methods, metrics, and policies) for the stakeholders. While the insights may include the improvement of serious game design, stakeholders are much more interested in the ability of the 'tool' to develop skills, raise performance, and improve bottom-line – through the reduction of training cost and its impact on return of investment (see Kozlov & Reinhold, 2007; Loh, 2012a).

Even though, to date, few serious games have an assessment component, the concern for the lack of appropriate methods and metrics for performance measurement is not new (Michael & Chen, 2005). For example, Crookall (2010) asserted that serious games and simulations could add more values by incorporating appropriate debriefing tools for performance assessment and improvement. Seif El-Nasr, Drachen, and Canossa (2013) advocated turning user-generated data into game analytics for monetization. Loh, Sheng, and Ifenthaler (in press) explored various methodologies to measure, assess, and improve performance of training and learning in serious games in a new edited volume, entitled *Serious Games Analytics*. To better meet the needs of various stakeholders, one has to discover useful metrics (when none is available) for measuring human performance with serious games, identify strong predictors (from among many other weaker ones) for incorporation into methods of 'best-practice', and maximize the value of the analytics for return of investment and performance improvement.

Serious games analytics requires a two-step process. The first step is the collection of user-generated data to ascertain what has been *done* in the training environment, in order to extrapolate

* Corresponding author at: Virtual Environment Lab (V-LAB), Department of Curriculum & Instruction, Southern Illinois University, 625 Wham Drive, Carbondale, IL 62901-4610, USA. Tel.: +1 618 453 4206.

E-mail addresses: csloh@siu.edu (C.S. Loh), ysheng@siu.edu (Y. Sheng), henryryu@siu.edu (I-Hung Li).

[1] Address: Department of Counseling, Quantitative Methods, and Special Education, Southern Illinois University, 625 Wham Drive, Carbondale, IL 62901-4618, USA. Tel.: +1 618 453 6913.

[2] Address: Virtual Environment Lab (V-LAB), Department of Curriculum & Instruction, Southern Illinois University, 625 Wham Drive, Carbondale, IL 62901-4610, USA. Tel.: +1 618 503 0188.

http://dx.doi.org/10.1016/j.chb.2015.02.053
0747-5632/Published by Elsevier Ltd.

what has been *learned*. Bellotti, Kapralos, Lee, Moreno-Ger, and Berta (2013) reported that the most prevalent methods found in the serious games literature are user surveys and pretest–posttest methods. After analyzing more than 510 data collection techniques used before, after and during digital games, Smith, Blackmore, and Nesbitt (in press) reported that the majority of techniques (before and after) were questionnaires, followed by some kind of test (38% and 31%, respectively). In contrast, these two methods were inefficient for 'during game' data collection (0% and 1.5%, respectively). This is because *ex situ* methods (such as questionnaires, surveys, and pre/post-tests) are less objective compared to *in situ* user-generated data for serious games analysis: user-surveys are self-reported data (Fan et al., 2006; Hoskin, 2012) and pre/posttest treats games as an impenetrable Black Box (Loh, 2012b). In comparison, *in situ* data collection methods, using telemetry, trace players' actions and behaviors from within the gaming habitat itself. User interactions with (predetermined) game events – e.g., objectives met, number of enemies killed, navigational route(s) taken, items discovered, etc., are directly recorded via 'event listener' functions *in situ* and then stored as game logs (on the local machine) or in an online database (on a remote machine). *In situ* data are relatively free of 'noise' caused by human data-input errors, in addition to being more objective by nature.

While a game log (being a plain text file) is the easiest to produce and is probably the most common option in game analytics research (Drachen, Thurau, Togelius, Yannakakis, & Bauckhage, 2013), its usefulness is limited to *post hoc* (after action) report because a log needs to be further processed before analysis. A better alternative is to capture and store the user-generated data using an online database (Loh, Anantachai, Byun, & Lenox, 2007; Zoeller, 2013) – especially if an online gaming architecture is already in play. The advantage of online database over game log is that the data are already optimized, and are therefore, immediately available for real-time analytics and *ad hoc* reporting (Ellis, 2014). The concept of real-time analytics of serious games by combining telemetry and online database to facilitate *ad hoc* reporting is not new, viz. *Information Trails* (Loh, 2006). The *Information Trails* communicates its analytics in both *ad hoc* (real-time) and *post hoc* analysis by way of a visualization component, viz. *Performance Tracing Report Assistant* (or PeTRA; Loh, 2012a, 2012b).

Once the user-generated gameplay data become available, the second step in the serious games analytics process is to mine or analyze the data for any hidden patterns therein – e.g., via statistical or machine learning methods and pattern recognition techniques. As there can be literally thousands of predictors available (depending on the size of the data), the key is to weed out the weak predictors by identifying the strong ones. The goal is to (a) identify features in serious games (design) for the measurement of players' skills, abilities, and knowledge, (b) assess players' performance as evidence of usefulness of serious games, and (c) facilitate the formulation of new policies/insights for performance improvement in the human trainee/learners and the serious games for training/learning.

2. Related studies

2.1. Expert–novice differences

A clear understanding of the differences between experts and novices can help us understand how knowledge is acquired, target the differences for (re)training and improvement, and teach new (systemic) skills in a variety of situations: e.g., robotic surgery, aviation, sports, music performance, strategic thinking, disaster preparation, behavior recognition, and many others. Research about behavioral and cognitive differences between skilled individuals and novices – or *expertise* in general, began in earnest around the turn of the 20th century (e.g., Bryan & Harter, 1899; de Groot, 1978) and continued through modern days (Dreyfus & Dreyfus, 2005; Ericsson, Prietula, & Cokely, 2007; Hong & Liu, 2003; Ifenthaler, 2010; Rauterberg, 1995).

There is a general agreement that expertise is not an innate ability but the result of *deliberate practice* (Ericsson, Charness, Feltovich, & Hoffman, 2006), which enables task-performers to gradually hone their skills by overcoming the limitations in working memory and sequential processing. The cognitive-behavioral indicators of expert–novice performance differences have been demonstrated variously in: time-to-task-completion (Cappiello et al., 2011; Hornbæk, 2006), mental representations (Kozma & Russell, 1997; Wiedenbeck, Fix, & Scholtz, 1993), dynamic decision-making (Gonzalez, 2005; Gonzalez & Golenbock, 2003), gaze patterns (Law, Atkins, Kirkpatrick, & Lomax, 2004; Underwood, 2005), neural/perceptual responses (Mishra, Zinni, Bavelier, & Hillyard, 2011), and numerous others. Besides qualitative modeling (Rauterberg, 1995), expert–novice performance differences in empirical studies can be coded using observable, quantifiable, and differentiable behavioral traits – such as counting the number, types, and range of specific actions performed (Boot, Blakely, & Simons, 2011; Boot, Kramer, Simons, Fabiani, & Gratton, 2008; Loh et al., 2007), or errors performed (Frese & Zapf, 1994) within a certain time frame.

More recently, researchers have begun coding behavioral data using action sequences – i.e., the chronological order of actions performed to fulfil specific task(s) under observation. The advent of high-speed, high-resolution digital computing technologies is one important factor for the renewed interest in action sequence studies: e.g., digital camera for kinesiology and sports learning (e.g., Williams, 2000; Williams & Ericsson, 2005), Internet for Information Scents (Chi, Pirolli, Chen, & Pitkow, 2001), Web searching strategies (Awad & Khalil, 2012; Hölscher & Strube, 2000), and digital games for game-based problem-solving (Loh & Sheng, 2013; Weber & Mateas, 2009). As the studies of Internet and digital games involve human perceptual-cognitive interaction with computers, such research is also within the purview of perceptual learning and human–computer interaction.

2.2. Expertise and skill acquisition

The literature on expertise training indicated that it was indeed possible to train *perceptual expertise* in a laboratory environment (e.g., Gauthier, Williams, Tarr, & Tanaka, 1998; Krigolson, Pierce, Holroyd, & Tanaka, 2009; Tanaka, Curran, & Sheinberg, 2005). The notion that expertise can be trained in a laboratory setting (i.e., outside the actual training context) has an important implication for training. If expertise can indeed be trained, would it be possible to refine the training sequences, and thus, shorten the extent of deliberate practice required? Further, can artificial training systems (such as serious games and simulations) be created to help train or produce more experts?

In industries that require highly complex skills (e.g., aviation and surgery), the ability to produce more experts in a shorter timeframe can have greater implications, as it is both cost- and life-saving. As military combat, medical surgery, and other human services become increasingly reliant on serious games for skill training (e.g., Rosenberg, Landsittel, & Averch, 2005; Sabri et al., 2010), serious games analytics (Loh et al., in press) and insights (Loh & Sheng, in press) are set to become a necessity in establishing new policies for the training of skilled performers to expertise.

The phases of skill acquisition were first discussed by Fitts and Posner (1967) (as described by Beilock and Carr (2004)), using a 3-stage model. Their model is composed of three phases: (1) a Cognitive phase where novices would consciously follow the task