ACCEPTED BY INTERNATIONAL JOURNAL OF INDUSTRIAL ERGONOMICS (2016)

The Five Key Questions of Human Performance Modeling

Changxu Wu

State University of New York at Buffalo

Abstract

Via building computational (typically mathematical and computer simulation) models, human performance modeling (HPM) quantifies, predicts, and maximizes human performance, human-machine system productivity and safety. This paper describes and summarizes the five key questions of human performance modeling: 1) Why we build models of human performance; 2) What the expectations of a good human performance model are; 3) What the procedures and requirements in building and verifying a human performance model are; 4) How we integrate a human performance model with system design; and 5) What the possible future directions of human performance modeling research are. Recent and classic HPM findings are addressed in the five questions to provide new thinking in HPM's motivations, expectations, procedures, system integration and future directions.

Keywords: Human Performance Modeling

1. Introduction

This article elaborates on the five key questions in HPM, describing its motivations, expectations, procedures, system integration and future applications. Since the area of human factors and ergonomics is quite large, this article mainly focuses on the modeling of cognitive-related human performance (e.g., cognition and motor performance under the control of cognition).

2. Q1. Why do We Build Models of Human Performance?

In science, besides unifying many scattered findings from empirical studies (Card, Moran, & Newell, 1983), models of human performance provide a systematic and computational understanding of the mechanisms of human behavior. In many experimental studies, verbal descriptions (sometimes conceptual models) of mechanisms are very important; however, they cannot make accurate prediction of human performance. Moreover, since the human cognitive and motor system is very complex, verbal descriptions in many cases may not quantify these complex relationships. Computational models of human performance can solve these problems. In addition, models can also guide researchers in data collection and provide researchers with a baseline against which to measure human performance (Sinclair & Drury, 1979).

In engineering, models of human performance can help system designers save significant amount of time and cost in running experiments, and also be integrated into the intelligent/smart systems directly to improve system safety and human performance and/or prevent accidents. The ability to predict human behavior means that, in many cases, accidents are prevented and errors are minimized to improve system safety and efficiency. For example, a human performance model can predict speeding behavior of a driver a few seconds before the actual speeding behavior occurs (Zhao, Wu, & Qiao, 2013). Once it was embedded in an intelligent system, the system could send pre-speeding warning to drivers to prevent traffic accidents before they occurred (Zhao & Wu, 2013).

3. Q2. What are the Expectations of a Good Human Performance Model?

The expectations of a good human performance model can be summarized into the following aspects: Mechanisms, Usefulness, Robustness and Generality, and Simplicity (Called as MURGS expectations in HPM).

Mechanisms ("Does this model address the mechanisms of human performance?"): As we discussed in the motivation of human performance modeling, a good model should quantify the relationship between the model's input and output based on the human cognitive and/or motor systems' mechanisms; otherwise, the model may be downgraded to a "black box" model. This issue is related to the *difference between top-down (theory-driven) human performance models and bottom-up (data-driven) models*—including artificial intelligence models (e.g., artificial neural network (ANN) models) and statistical models), since most bottom-up (data driven) models can relatively easily capture the relationships between model's input and its output (data to be modeled) via model training; however, usually bottom-up (data-driven) models do not quantify the fundamental mechanisms of the human or human-machine systems, or their modeling mechanisms are different from the mechanisms of human cognition and motor system (they have their own sets of modeling/quantification rules). Moreover, due to the lack of the top-down understanding of the mechanisms of human or human-machine systems, bottom-up (data driven) models may over-fit one data set with extensive training for that data set, but under-fit a new data set, leading to their problems in robustness and generality (in other words, leading to the "missing the forest" problem).

Usefulness ("Can this model, once built and verified by the data, improve real-world system performance/safety/efficiency?"): Different from cognitive modeling, such as the work of Isbel &

Mahar (2015), that focuses more on the mechanisms and human behavior in lab settings, the emphasis of human performance modeling is more on the human performance and safety in practice and real-world settings. Accordingly, the first expectation of a good human performance model is that its prediction should be directly related to human performance and be useful in real-world system design to improve the performance, safety, and efficiency of human operators and/or the human-machine system as a whole in real-world settings.

Robustness and Generality ("Can this model predict multiple experimental results without over- or under-fitting?"): This expectation of a good model includes two parts: a) Avoid over-fitting or under-fitting; and b) Verification by multiple empirical studies. A good model should not only avoid under-fitting the data (e.g., R square between the model's prediction and experimental data is below 0.5), but also avoid over-fitting the data. Over-fitting usually means a perfect match between the prediction and the experimental data from one study but a poor match between the prediction and the experimental data from another study (See detail discussions of the model over-fitting and under-fitting issues in the work of Lewandowsky & Farrell (2010)). For example, given the same root-mean-square (RMS) of the two models (A and B), Model A verified by two experiments (R square for Experiment 1=0.75 and 0.71 for Experiment 2) is more robust than Model B whose R square for Experiment 1=1 (over-fitting) and 0.46 (under-fitting) for Experiment 2, even if their averaged R square is the same (0.73).

Simplicity ("Is this the simplest model for making a useful and a robust prediction based on human performance mechanisms?"): This expectation is also very important in evaluating a human performance model. This simplicity rule is the same as the parsimonious rule in mathematical and simulation modeling in general: a simpler model is better than a complex model as long as they achieve the same level of functionalities. Moreover, mathematical models are preferred in general than simulation models unless NP-Hard or no analytic solution problem has been encountered by mathematical models (Bank, 2000). The simplicity in HPM is defined as the number of free parameters (The parameters of a model whose values are estimated from the data to be modeled to maximally align the model's prediction) (Lewandowsky & Farrell, 2010), the format and structure of equations if it is a mathematical model, and the number of lines of codes in general if it is a simulation model. For example, a linear model is better than a non-linear model with the same number of parameters as long as both models meet the other three expectations at the same level. Another way to compare the simplicity of different models is to calculate their AIC (Akaike

Information Criterion) which considers the number of free parameters (Busemeyer, 2000); however, AIC does not consider structures of equations or the number of lines of computer simulation codes.

4. Q3. What are the Procedures and Requirements in Building and Verifying a Human Performance Model?

Depending on the availability of the data and interests of the modeler (the person who builds a model), we summarize the three different approaches to carry out the human performance modeling work.

Approach 1 (***Conceptual/Existing Model or Theory→Model→ Verify the Model by Other People Later***) In situations that there is no data available or modelers are not able to conduct experiments to verify the model, researchers can still propose/build model without its verification from data. A classic example of modeling work is Einstein's Relativity Theory which was proposed based on theory without experimental data to verify the model's predictions directly at the time when the model was proposed (Einstein, 1905). After a few decades when technologies were feasible to carry out the experiments, the Relativity Theory was eventually verified by the experimental data directly (Hafele & Keating, 1972). We actually think that this is one of acceptable ways of modeling to avoid a modeling problem—If the modeler did have data prior to building a model, he/she could learn what patterns exist in the data during the modeling process and change the model to fit that data, consciously or subconsciously.

In situations that data are available (either from existing published work or from a modeler's own experiments), we typically regard these modeling processes as a mathematical statement proof process (e.g., prove the "$a^2+b^2=c^2$" Pythagorean Theorem). This is because the modeler receives data (prediction of the model) before the model is built (although the published modeling work is usually written in reverse order, presenting the model first and model verification with data second). Therefore, a modeler ***should*** clearly provide step-by-step details outlining how his/her model reaches the final prediction (the model's prediction will "definitely" be verified by the data, otherwise the modeler will not even submit this modeling work). If it is a mathematical model of human performance, a modeler should list all of the model derivation steps clearly from the model input to the model output (prediction of the data), without skipping any important steps; If it is a computer simulation model, a modeler should list and ***describe*** the meaning of all the code in the simulation (just listing the computer code at the end of the paper or putting them on a website may not be enough since it is very difficult for a reviewer/reader to understand the code without the author's descriptions), to ensure that there are no codes purposely added in the simulation codes to make the model fit the data.

Approach 2 (***Conceptual/Existing Model or Theory→Model 1→Verify Model (Needs Improvement)→Model 2 etc.***)

Some modeling work has treated modeling as an iterative process, such as EPIC's modeling study (Kieras & Meyer, 1997). Based on conceptual models, theory, or existing models, researchers built a relatively simple model first, verifying the model with the data; during the verification process, researchers learned that the simpler model is missing some important component(s), and then they improved the model to improve its prediction. The improved model was then verified with the data again.

Approach 3 (***Conceptual/Existing Model or Theory→Model→Verify the Model***)

Some studies, such as (Du, Shi, & Yuan, 2007), in human performance modeling area appear to have skipped the iterative process in human performance modeling (judged by the work presented in their papers), proposing models immediately followed by their verification with the data. When either Approach 2 or 3 is used, a modeler should clearly show how their models reach the prediction step by step since they receive the data before the development of the model. In addition, a modeler should also report the number of free parameters used in the modeling process. Approach 3 also includes another procedure that researchers proposed a new model and then develop hypothesis to collect data based on this new model, where newly collected data will be further used to verify the hypothesis and the new model (Takahiro Wada, Konno, Fujisawa, & Doi, 2012; T. Wada & Yoshida, Accepted).

In the **verification stage** of a human performance model, we suggest each modeling work reports R square and Root-Mean-Square (RMS), since R square can reflect how the model's prediction captures the changes of the patterns of experimental data and RMS reflects the absolute difference (i.e., magnitude of prediction error) between the model's prediction and experimental data (See Fig 1. The Y axis is usually an index of human performance (e.g., task completion time; X axis can be an independent variable in an experiment (e.g., task difficulty level), time, or other variables as an input of the model). However, there are two situations that R square cannot be used

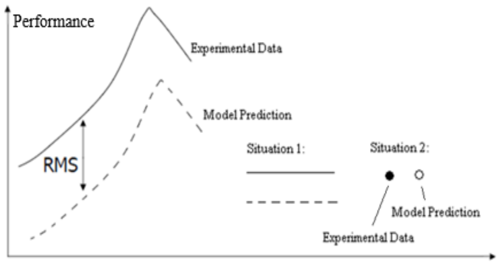

Figure 1. R Square and RMS

as an index of model verifications: Situation 1: When both the model's prediction and experimental data are in parallel with the X axis: R square cannot be calculated in this case since the denominator of the R square's equation is 0 (See Fig 1). Situation 2: There is only one data point in the experimental data (e.g., error rate of the human operator) and we need at least 2 data points in the experimental data to calculate R square between the model's prediction and experimental data. In these situations, we can use the percentage of estimation error (John, 1996) to verify a model: (experimental data - model's prediction) / experimental data×100% or |experimental data - model's prediction|/ experimental data×100%. In addition, some researchers performed statistical analysis (e.g., ANOVA or *t* test) to check whether there is a significant difference between experimental data and a model's predictions (Liu, 2005).

5. Q4. How do we Integrate a Human Performance Model with System Design?

Ideally speaking, a good human performance modeling work should include three major steps: Building a model (including extending an existing model), verifying the model's prediction with data, and applying the model in system design (sometimes referred to as "playing" a model). In many published modeling studies, the third step is optional; however, we expect modeling work should at least describe this third step verbally since we are building useful human performance model in practice (See Pan and Bolton's paper in this special issue). Besides the traditional approach of prediction of human performance, there are at least two additional approaches that we can "play" a model after it is built and verified by experimental data.

Playing Approach 1: Optimizing a model's input to maximize or minimize its output

The goal of human performance modeling is to maximize the safety, efficiency, and (or) performance of the human-machine system. Accordingly, we can build objective functions to maximize or minimize a model's output (human performance including error rate, task completion time, workload etc.) by treating the input of the model as decision variables. For example, after a mathematical model of human performance in a numerical typing task was built, we can derive a set of objective functions to maximize a human operator's typing performance (model's output) and treat the key sizes and gaps among keys (model's input) as decision variables. The solutions of the objective functions will determine the optimal values of the key sizes and gaps among keys, which can be directly used in system design (Lin & Wu, 2012).

Playing Approach 2: Prediction of human performance can be an important input of intelligent system design to achieve safety and overall system performance.

The output of a human performance model can serve as an important input for various intelligent systems. For example, after a lane-change model of human drivers was developed (Salvucci, In press), we can predict when a driver is going to change lanes before that actual behavior occurs and then design intelligent systems to send pre-warnings to the current and other drivers to prevent traffic accidents. Another example in transportation safety is speeding prediction and prevention. Once a human performance model was able to predict the over-speeding behavior of a driver a few seconds before that behavior actually happens, an intelligent pre-warning system was designed to receive this output from the model and send pre-warnings to the driver (Zhao et al., 2013). An experimental study validated that the model-based intelligent pre-warning system was able to achieve significantly better safety benefits (e.g., smaller magnitude of speeding) compared to traditional post-warning system (warning is sent to a driver after speeding is detected) (Zhao & Wu, 2013).

6. Q5. What are the Possible Future Directions of Human Performance Modeling Research?

Allen Newell's dream for building unified models that can mimic almost every activity of human beings (e.g., driving a car and daydreaming) has been the major direction of human performance modeling research for a long time (Newell, 1973). Moreover, there are several new directions of human performance modeling for modelers to consider:

1) Solving the Next-Moment Prediction Challenge by Integrating HPM with Data-Driven Models/Methods

Prediction of an individual's behavior in the next moment (e.g., a few seconds, minutes or hours in advance depending on dynamic properties of the human-machine system) under a specific situation is one of the most important predictions in practice, so that intelligent systems can predict and interact with human operator(s) almost in real time and interrupt problem behaviors. We call this the ***next-moment prediction challenge***. Given the dynamic changes of human operators, their tasks, and environment, it is very hard for top-down models to solve this challenge since the data are only served as a way to verify the model and lots of information in the data are underutilized by the top-down human performance models. However, data-driven approaches (e.g., data mining methods) have the limitations in lacking a top-down theoretical understanding of the mechanisms, leading to the "missing the forest" problem. Accordingly, the integration between the top-down HPM and data-driven models/methods may solve the ***next-moment prediction challenge*** and lead to an important direction for HPM (Lin, Wu, & Chaovalitwongse, 2015).

2) Integrating with System Design

Usefulness and system design integration are the two features of HPM that makes it different from cognitive modeling. Future work in HPM should emphasize how a human performance model is able to assist system design to improve human performance. Compared with experimental findings, HPM can be embedded in intelligent systems directly to predict and optimize human performance (including workload) when the context of tasks and human information processing capacities change (Wu, Tsimhoni, & Liu, 2007; Zhao & Wu, 2013).

3) Driven by new modeling theories: Future HPM will benefit from the advances of other modeling theories (e.g., Chaos Theory (Dafilis, Frascoli, Cadusch, & Liley, 2013; Jin & Chen, 2016) and Theory Of Everything (TOE) (Weinberg, 1992)). New theories will not only provide HPM modelers with new modeling approaches but also shed light on prediction of human performance and behavior at their stochastic aspects. Even though Theory Of Everything (TOE) is currently focused on our physical world, we will not rule out the possibility of incorporating HPM in TOE (Vimal, 2010) since the human cognition system is an important part of the world and the human brain is still a physical system in essence.

Acknowledgement

We appreciate the support from National Science Foundation for this work. We also appreciate the valuable suggestions from the anonymous reviewers in improving this work.

Reference

Bank, A. (2000). *Discrete Event Simulation, Taylor & Francis Group*.

Busemeyer, J. R. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44*, 171-189.

Card, S., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hinsdale, NJ: Lawrence Erlbaum.

Dafilis, M. P., Frascoli, F., Cadusch, P. J., & Liley, D. T. J. (2013). Four dimensional chaos and intermittency in a mesoscopic model of the electroencephalogram. *Chaos, 23*(2). doi:10.1063/1.4804176

Du, J., Shi, H., & Yuan, X. (2007). Modeling of human's pointing movement on the effect of target position *Digital Human Modeling*: Springer-Verlag Berlin Heidelberg.

Einstein, A. (1905). On the Electrodynamics of Moving Bodies. *Annal Physik, 17*, 891-921.

Hafele, J. C., & Keating, R. E. (1972). Around-the-world atomic clocks: observed relativistic time. *Science, 177*, 168-170.

Isbel, B., & Mahar, D. (2015). Cognitive mechanisms of mindfulness: A test of current models. *Consciousness and Cognition, 38*, 50-59. doi:10.1016/j.concog.2015.10.005

Jin, W., & Chen, F. (2016). Topological chaos of the spatial prisoner′s dilemma game on regular networks. *Journal of theoretical biology, 391*, 43-50. doi:10.1016/j.jtbi.2015.11.016

John, B. E. (1996). TYPIST: A theory of performance in skilled typing. *Human-Computer Interaction, 11*(4), 321-355.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 12*(4), 391-438.

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: principles and practice, Taylor & Francis Group.*

Lin, C., & Wu, C. (2012). Mathematically modelling the effects of pacing, finger strategies and urgency on numerical typing performance with queuing network model human processor. *Ergonomics, 55*(10), 1180-1204. doi:10.1080/00140139.2012.697583

Lin, C., Wu, C., & Chaovalitwongse, W. (2015). Integrating Human Behavior Modeling and Data Mining Techniques to Predict Human Errors in Numerical Typing. *IEEE Transactions on Human-Machine Systems, 45*(1), 39-49.

Liu, Y., Feyen, R., and Tsimhoni, O. (2005). Queuing Network-Model Human Processor (QN-MHP): A computational architecture for multi-task performance in human-machine systems. *ACM Transaction on Human Computer Interaction, In Press*.

Newell, A. (1973). You can not play 20 questions with nature and win. Projective comments on the papers of the symposium. In W. G. Chase (Ed.), *Visual Information processing*. Washington, D.C.: Academic Press.

Salvucci, D. D. (In press). Modeling driver behavior in a cognitive architecture. *Human Factors*.

Sinclair, M. A., & Drury, C. G. (1979). On mathematical modeling in ergonomics. *Applied Ergonomics, 10*(4), 225-234.

Vimal, R. L. P. (2010). Towards a theory of everything: Part I Introduction of consciousness in electromagnetic theory, special and general theory of relativity. *Neuroquantology, 8*(2), 206-230.

Wada, T., Konno, H., Fujisawa, S., & Doi, S. i. (2012). Can passengers' active head tilt decrease the severity of carsickness? effect of head tilt on severity of motion sickness in a lateral acceleration environment. *Human Factors, 54*(2), 226-234. doi:10.1177/0018720812436584

Wada, T., & Yoshida, K. (Accepted). Effect of passengers' active head tilt and opening/closure of eyes on motion sickness in lateral acceleration environmentment of cars. *Ergonomics*.

Weinberg, S. (1992). *Dreams of a Final Theory: The scientist's search for the ultimate laws of nature.*: Knopf Doubleday Publishing Group.

Wu, C., Tsimhoni, O., & Liu, Y. (2007). *Development of an adaptive workload management system using the Queueing Network-Model Human Processor.* Paper presented at the 51st Annual Meeting of the Human Factors and Ergonomics Society, HFES 2007, October 1, 2007 - October 5, 2007, Baltimore, MD, United states.

Zhao, G., & Wu, C. (2013). Effectiveness and acceptance of the intelligent speeding prediction system (ISPS). *Accident Analysis and Prevention, 52*, 19-28. doi:10.1016/j.aap.2012.12.013

Zhao, G., Wu, C., & Qiao, C. (2013). A Mathematical Model for the Prediction of Speeding with its Validation. *IEEE Transactions on Intelligent Transportation Systems, 14*(2), 828-836. doi:10.1109/TITS.2013.2257757