Bayesian Cognitive State Modeling For Adaptive Serious Games

Alexander Streicher[1] and Michael Aydinbas[2]

[1] Fraunhofer IOSB, Karlsruhe, Germany
alexander.streicher@iosb.fraunhofer.de
[2] TU Darmstadt, Germany
michael.aydinbas@gmail.com

Abstract. Bayesian modeling of cognitive state is one possible approach to user modeling for use with adaptivity in serious games. Adaptive educational serious games try to keep learners engaged - to keep them in the so called “Flow” channel, i.e., in the right balance between being challenged and entertained. The challenge is to intervene adaptively at the right time. The research question is when to actually adapt, and how to find quantifiable metrics for that. One way to achieve this is to model the users’ cognitive state and to adapt to high or low cognitive load, e.g., to apply dynamic difficulty adjustments. Our user modeling approach is based on Hierarchical Bayesian Models (HBM) which are suitable for drawing conclusions about the learner’s cognitive state inferred from observable variables. An important aspect is that the approach considers activity stream data such as from the Experience API (xAPI) protocol as input to achieve high interoperability and eased applicability. An evaluation with synthetic data for different user group types shows the feasibility of the approach. The model can explain differences between subjects, between subject groups and between different latent variables such as cognitive load or mental working memory capacity.

Keywords: user modeling, cognitive modeling, learner state, adaptivity, serious games, Bayesian inference

1 Introduction

A key question for an adaptive education system (AES [15]) is when adaptability should occur, that is, at what point a learner should be best supported. This also concerns serious games, i.e., (digital) games with the characterizing goal to educate and not just to entertain [5,9]. The challenge is to infer the needs of the users by observing how the users interact with the systems. AES could try to determine the cognitive states of the users to trigger the right adaptive responses [14,20]. For example, an adaptive educational system could react when attention decreases, cognitive load increases [14], or when the user seems to be in a repetitive cycle with no real observable progress, or when there are signs of forgetting. Cognitive modeling can provide answers to these questions by focusing on user modeling [14,20]. The principal part of an adaptive system for educational serious games is the user or student model [19] which can include information about the learners' current cognitive states, e.g., cognitive load or stress level, motivation, attention, etc. The model's estimate of the learner's current cognitive state can be used to derive and suggest adaptive measures to help navigating the player through the game on a smooth and engaging path.

The research question of this work is how to create a cognitive user model which allows to estimate latent cognitive variables such as cognitive load, perceived difficulty, or prior knowledge level. Subsequently, AES can use such a model to determine the best time for an adaptive response.

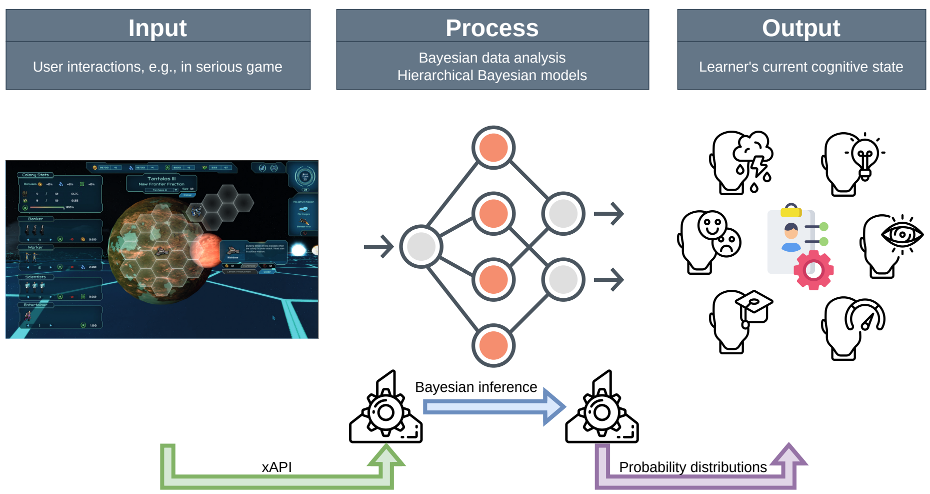

Fig. 1: Input-Processing-Output (IPO) concept for applying generative Bayes models to user models; output are probabilities about user cognitive states.

The contribution of this work is a concept for a cognitive user model which is based on *Hierarchical Bayesian Models* (HBM). The model makes use of the *Cognitive Load Theory* (CLT) [17] to describe the relationship between the characteristics of the learner and the characteristics of the learning material. The model has been validated numerically using synthetic data. A detailed model comparison was conducted to obtain the best model. The final model is able to explain the three observable variables task success, mission score and mission time. In our technical evaluation with synthetic data it can accurately predict those variables with an accuracy of over 90% (mean absolute deviation).

As depicted in Fig. 1, input to the model are activity stream data formatted according to the *Experience API* (xAPI) protocol [1]. This allows for increased interoperability and applicability to other application domains. The xAPI specifies how to capture user interaction events as a stream of activities. Every activity is based on an actor-verb-object triple structure, with additional information such as timestamps, context data or the result outcome of an activity, e.g., an assessment result score. This leads to a machine- and human-readable stream of activities. We make use of xAPI as input for our modeling approach.

The model's output are probability distributions that allow for inferences about the model's parameters and provide the uncertainty associated with those inferences.

To the best of our knowledge, this is the first application of a HBM for realizing cognitive user models for adaptive serious games. Several issues with the final model have been identified and provide directions for further research.

2 Related Work

Seyderhelm et al. (2019) propose a Cognitive Adaptive Serious Game Framework (CASG-F), routed in the cognitive-affective theory of learning with media, that combines performance measures and cognitive load to adapt the in-game tasks [14]. They suggest using a real-time, virtual detection-response task embedded in serious games to measure cognitive load and provide an adaptation template for six different combinations of performance and cognitive load measures [14].

Conati et al. (2020) investigated the usage of interaction data as an information source to predict cognitive abilities [3]. In a user study, they compared the predictive performance for cognitive abilities using only interaction data, eye-tracking data and both interaction and eye-tracking data. The researched cognitive abilities were perceptual speed, visual working memory, spatial memory, visual scanning, and visualization literacy. To measure the cognitive abilities, the participants had to take a series of tests after completing the actual task. While eye-tracking data generated the most accurate predictions, results showed that interaction data can still outperform a majority-class baseline. Additionally, it was found that interaction data can predict several cognitive abilities with better accuracy at the very beginning of the task than eye-tracking data, which is valuable for delivering adaptation early in the task. Left click rate and time to first click were the top two predictors for all cognitive abilities, suggesting the importance of those two features for predicting cognitive abilities. Conati et al. (2020) concluded that adaptation for interactive visualizations tasks could be enabled using solely interaction data [3].

Hallifax et al. (2020) analyzed the effect of combining several learner models to guide the adaptation strategy. They showed that adaptation is more effective when tailored to both player type and motivation, which could improve the intrinsic motivation to engage with the content [8].

Tadlaoui et al. (2018) realized a probabilistic and dynamic learner model in adaptive hypermedia educational systems based on multi-entity Bayesian networks [18]. The model can represent the different actions that the learner can take during their learning path.

One classical example of using dynamic Bayesian networks to model the causes and effects of emotional reactions was given by Conati and Maclaren

(2009) [4]. Their diagnostic model targets affect and how emotions are caused by the users' appraisal in a given context, e.g., for goals or preferences.

Gelman et al. (2013) state that checking the model is crucial to statistical analysis [7]. Bayesian prior-to-posterior inferences assume the whole structure of a probability model and can yield misleading inferences when the model is poor [7]. Therefore, good Bayesian analysis should include at least some check of the adequacy of the model's fit to the data and the model's plausibility for the purposes for which the model will be used. The most common way to check the fit of a model is by performing a posterior predictive check — as it was also done here. Posterior predictive distribution denotes a probability distribution over possible values of future data [11].

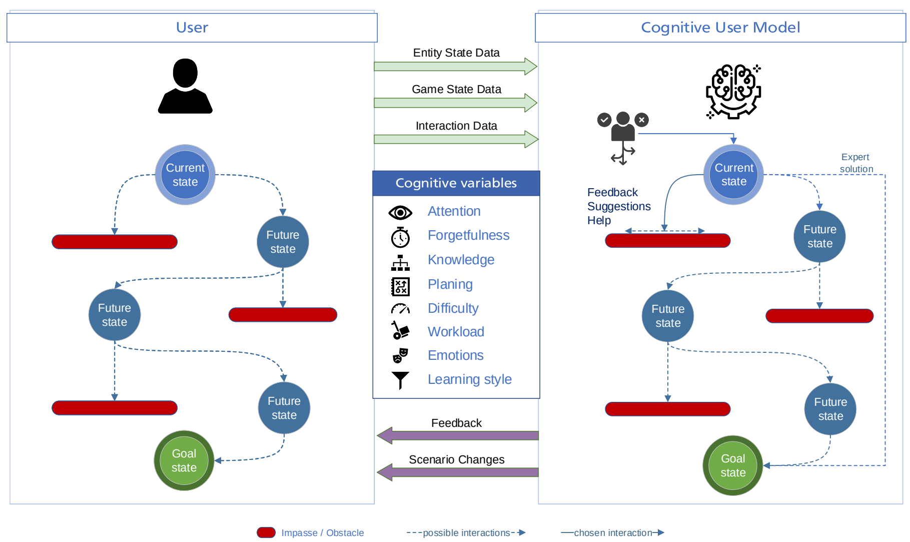

Fig. 2: Application principle for cognitive user models applied to (educational) serious games to estimate various cognitive variables as control input for adaptivity, e.g., adaptive feedback or scenario changes. (Image: Aydinbas 2019 [2])

3 Concept For a Cognitive User Model

A cognitive user or learner model [19] is a model that can make statements about the learner's cognitive state, that is, statements about their mental actions and processes that deal with knowledge acquisition and understanding. In the adaptive cycle by Shute et al. (2012) the learner model builds a connection between the captured user data and the presented learning material that is suitable for the learner [15]. The learner model should allow for a dynamic assessment of the learner's current cognitive state [3]. The model's knowledge about the learner can be leveraged to guide the player through the problem space towards a goal state while avoiding states that are detrimental for the player (e.g., [8,18]).

Fig. 2 depicts the idea how to apply cognitive modeling to serious games [2]. On the left is an outline of the user's cognitive states and how it develops while playing a game towards a goal state. As shown on the right, cognitive modeling tries to simulate that with a cognitive user model — the model tries to estimate various cognitive variables as input to adaptive control, e.g., adaptive feedback or scenario changes. This again results in an interaction between the user and the AES (as in a closed-loop feedback adaptive control).

3.1 Observable Variables

To work with Bayesian models, it is required to define the relevant observable data, the involved measurement scales of the data and the definition of variables that are to be predicted and variables that are predictors. For transferability reasons, and without loss of generality, we selected a variable set as found in most serious games. Three categories of observable variables were identified: general performance measures, domain-specific measures, and game-specific measures (Table 1). This can act as a framework for observable variables in serious games.

Table 1: Examples of xAPI observable variables in serious games.

Name	Level	Variable	Type	Unit	Domain
task success	performance	k	binary	-	$\{0,1\}$
mission score	performance	s	discrete	-	$\mathbb{N}_0$
mission time	performance	t	continuous	minutes/seconds	$\mathbb{R}_{\geq 0}$
required rounds	domain	n_{rnd}	discrete	-	$\mathbb{N}$
required hints	domain	n_{hnt}	discrete	-	$\mathbb{N}_0$
location changes	domain	n_{loc}	discrete	-	$\mathbb{N}_0$
dialogues	domain	n_{dia}	discrete	-	$\mathbb{N}_0$
detours	game	n_{det}	discrete	-	$\mathbb{N}_0$

General performance measures are mostly domain-independent and can be measured in any application. Games typically report on these measures in an assessment stage, e.g., at the end of a mission or task.

`Task success` k describes the success of a player for a given task and is either true or false. As an example, in an image interpretation serious game, a task or objective is the correct deployment of an imaging sensor, which can fail if the player has not correctly considered the actual weather conditions or has exceeded the number of rounds available for the mission.

`Mission score` s is the overall score of the player for the given mission, but is also applicable to any game and any learning task because it is a simple accuracy

measure that reflects the number of errors the player made. As an example, this number lies between zero and the maximal number of items the player had to report for a given mission, and typically it is normalized in $[0; 1]$ to be comparable across missions and across different applications.
`Mission time` t is the amount of time it took the player to finish the mission, but is applicable to any domain or learning task by measuring the time the learner needs to finish a given task.

Domain measures are domain-dependent, but applicable to any similar game of that domain that supports this feature. For example, strategy games use resources of some kind and turn-based games have a measure of rounds. Game measures are the most specific of all variables because they only apply to a particular game and are normally not transferable to other games or applications without reinterpretation. One example of a game measure variable might be `detours` which captures the number of detours the player has taken during the mission compared to an ideal solution where the player knows exactly what the next steps are and where to go.

3.2 Cognitive User Models For Serious Games

Instead of cognitive modeling frameworks such as Soar 9 [10], which demands a lot of domain and task modeling effort but cannot directly offer inferences about the cognitive state of the learner, we decided to choose a different approach: to build a probabilistic statistical model of the data that only specifies what is really needed for the user modeling task. The Bayesian modeling approach allows for maximal control and flexibility as the model can be as general or as complex as needed. Bayesian models are directly built in a way to infer the state of latent, non-observable variables from observable variables.

After the user has finished a game (or any other learning) session, either successfully or with failure, the collected data is analyzed by the HBM. The model takes predefined prior distributions for each model parameter and the observations as input and calculates (via Bayesian inference) the posterior distributions for all parameters. The latent parameters are variables that can be interpreted as cognitive variables such as motivation or the perceived difficulty.

Because the model produces posterior distributions, which are probability densities, the model gives directly interpretable probabilities as output. It is straight forward to derive point estimates such as mean and mode values as well as highest density intervals from the posterior distributions. At the end of each session, the model can infer the learner's current cognitive state and make statements about their current value with an estimate of uncertainty. Because the posterior is a probability density, it can be reused in another run of Bayesian inference as the prior, replacing the old, non-informative prior with the model's latest belief according to the observed data. This is identical to the prediction-correction step in Bayesian filtering methods, and the principle is depicted in Fig. 3. During each mission the player executes a series of actions, which takes time t. At the end of each mission all user interactions are send to the cognitive user model framework in form of xAPI statements. The framework is given prior distributions for each model parameter and observable data as input and produces, via Bayesian inference, posterior distributions for each model parameter. This posterior distribution is the base for the computation of point estimates. The posterior of one mission can also serve as the prior for another mission.

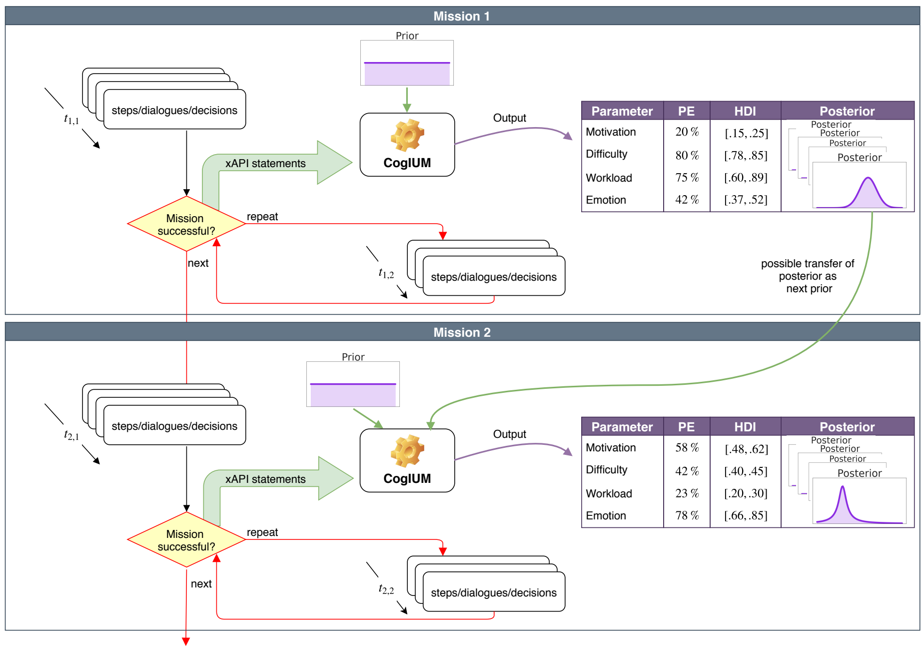

Fig. 3: Realizing a “Cognitive Intelligent User Model” (CogIUM) with HBM. After each round, task or mission, the model’s inferences can be used to adapt the game according to the learner’s needs, e.g., adjusting the difficulty level.

3.3 A Hierarchical Bayesian User Model

Our final model (Fig. 4) is a fully specified descriptive HBM that is able to model the three observable variables task success, mission score and mission time [2].

The model defines variables for two separate groups (depicted by rectangles): personal variables (indexed with p) that differ between subjects and conceptual variables (indexed with c) that differ between concepts. Concept in this context means a learning concept that the learner should acquire. In addition to group variables, which remain constant within the group, there are also variables that differ between subjects as well as between concepts (indexed with pc). This is the case for the observable variables, as they depend both on the current concept as well as on the current subject.

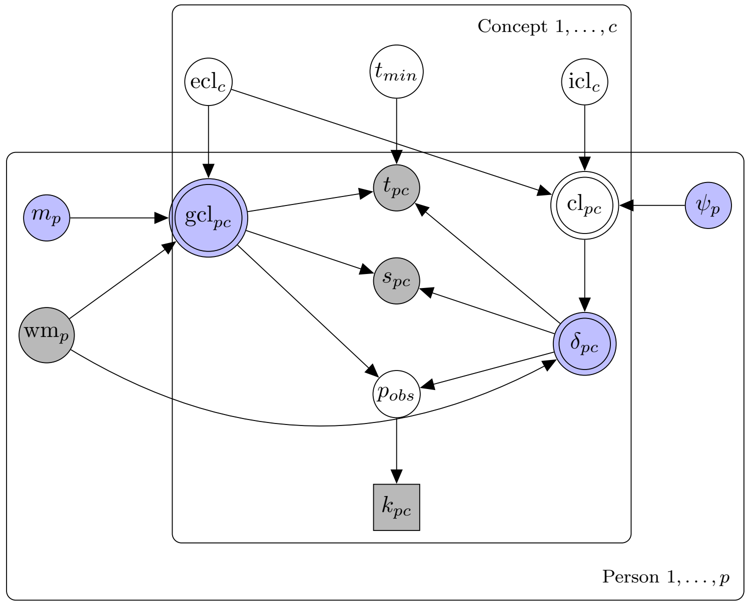

Fig. 4: Simplified graphical model of the best performing HBM (notation according to [6]). The fully specified descriptive model is described in [2] as $\mathcal{M}_4$.

The HBM models the cause of three observable variables: task success k_{pc}, mission score s_{pc}, and mission time t_{pc}. The latent variables (highlighted in blue) that serve as proxies for cognitive variables - which are the primary focus of the cognitive model - include motivation m_p, prior knowledge ψ_p, germane cognitive load glc_{pc}, and free working memory capacity δ_{pc}.

Besides the observable variables (Fig. 4, gray nodes), all other variables are considered latent and can be the target of inference. Especially, we are interested in variables that can be interpreted as cognitive variables and allow inference about the learner's cognitive state (Fig. 4, blue nodes).

The basic idea of this descriptive model is that cognitive load affects learner performance, which in turn has a direct impact on the observable variables. Higher cognitive load often means an increase in the number of errors, and thus a decrease in performance [12]. Cognitive load is purely defined by the sum of intrinsic and extraneous cognitive load [16]. Germane cognitive load is the proportion of working memory resources that deals with intrinsic cognitive load. Higher extraneous cognitive load means less germane cognitive load, which reduces learning and which can be measured by a decrease in the learner's performance [16]. For example, if the model observes that two different people perform differently when learning the same concept, it can attribute the differences to individual differences in cognitive load and free working memory capacity, which in turn are influenced by the learner's motivation and prior knowledge.

4 Evaluation

To validate the fitness and performance of our models, we generated 14 synthetic data sets [2]. Each data set contains plausible observations for a different number of subjects and missions (called concept). Participants were divided into groups of different "skills". There were three groups in total:

1. "good": a subject successfully completes the task, has a high mission score and a short mission time.
2. "average": a subject has a random chance to succeed in the task, and both has an average mission score and an average mission time.
3. "bad": a subject always fails the task, has a low mission score and a long mission time.

The data sets differ with respect to the number of subjects, the number of concepts, the number of groups, and how the subjects are divided among the groups. With the help of Bayesian inference, the model re-allocates credibility across parameter values to best explain the observations of each data set.

Comparing models is especially important since it is typically the case that more than one reasonable probability model can provide an adequate fit to the data in a scientific problem. A popular metric to quantify predictive performance is the *Mean Squared Error* (MSE) [13].

The best model was able to explain individual differences with more than 90% accuracy for the three observable variables task success, mission score and mission time.

In Table 2 we report the *Mean Absolute Deviation* (MAD), the *Mean of the Squared Deviations* (MSE), the average standard deviation (SD) and the accuracy (ACC). We define accuracy by comparing the MAD estimate with the maximum upper limit, in our case the maximum error of each variable: 1) task success is binary-valued $\{0, 1\}$ hence the maximum error is 1; 2) mission score is in $[0; 1]$ hence the maximum error is 1. Mission time can be any value greater zero, but we limited the session to 5 minutes. Table 2 shows that, on average, task success is reproduced 100% correctly for all subjects, mission score 98.9% and mission time 92.7%.

Table 2: Measures of predictive precision and predictive accuracy for the best model for a single simulated data set. P_i stands for the posterior predictive distribution of the i-th observation.

Observable Variable	MAD	SD_{MAD}	MSE	SD_{MSE}	ACC_{MAD}	SD
task success	0.0	0.0	-	-	1.0	0.46
mission score	0.011	0.011	0.001	0.001	0.989	0.107
mission time	0.366	0.203	0.427	0.445	0.927	1.963

$MAD = \frac{1}{n}\sum |y_i - \tilde{P}_i|$; $MSE = \frac{1}{n}\sum (y_i - \tilde{P}_i)^2$; $ACC = 1 - \frac{MAD}{\text{max. Error}}$; $SD = \frac{1}{n}\sum \sigma_{P_i}$

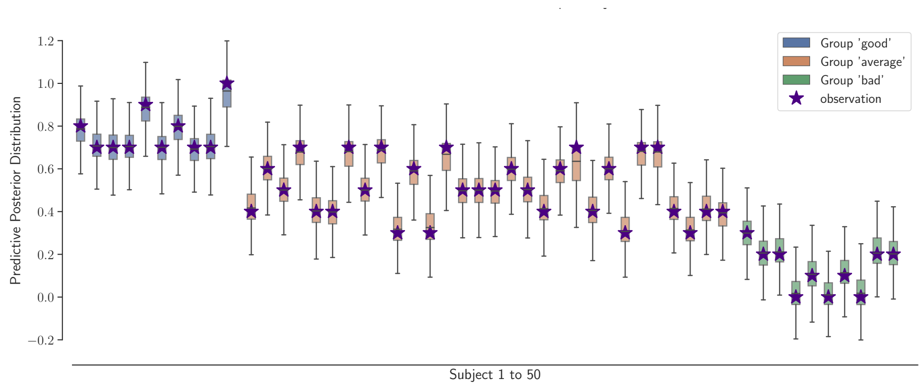

Fig. 5: Comparison between individual observations created by the model and the true observations (marked by a star) for mission score s_{pc}. The simulated data set includes 50 subjects, divided into three groups (good, average, bad). The model's predictions are based on samples drawn from the model's posterior predictive distribution.

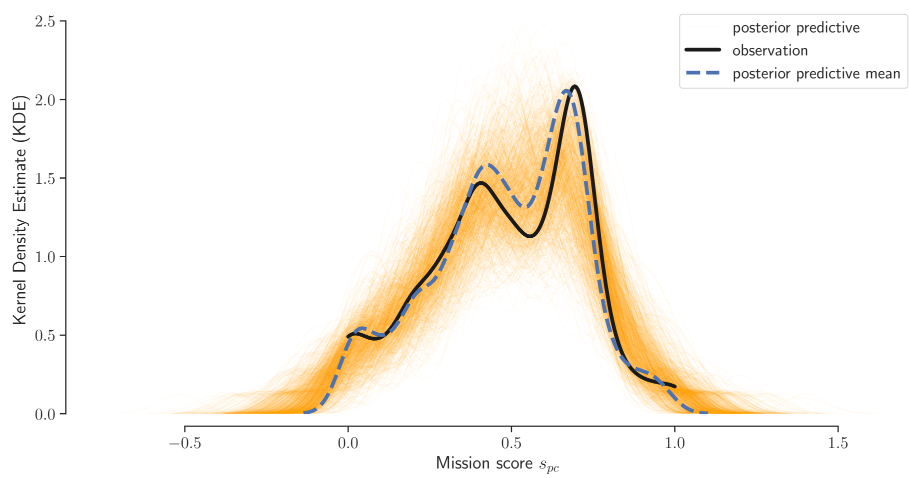

Fig. 6: Comparison between predicted observations (orange) by the model and the true observations (black) for mission score s_{pc}. Distributions were approximated by kernel density estimation for visualization. The predicted observations are based on samples drawn from the model's posterior predictive distribution. The mean posterior predictive distribution is plotted as a dashed blue line.

As an example, Fig. 5 and Fig. 6 show the results of a posterior predictive graphical check for the observable variable mission score for $n = 50$ subjects divided into the three distinct groups. The model is able to reproduce the observed individual scores accurately, and the posterior predictive mean curve is close to the observation curve.

The model works best for one concept. More than one concept becomes challenging for the final model and the model's performance decreases with an increased number of concepts. However, this behavior is expected as the HBM is primarily designed to explain individual differences and has not the ability to vary personal variables between concepts.

5 Conclusion & Outlook

Adaptive educational serious games try to continuously motivate players by dynamically adjusting the usage experience according to the users' cognitive states, e.g., according to the estimated free working memory or motivation. Cognitive user models can capture cognitive variables such as cognitive load, motivation, attention, etc. Adaptive systems can make use of such models to determine a good point in time when to actually adapt, e.g., a threshold on the cognitive load measure to control the displayed information scope, or to dynamically adjust the difficulty level. The research question is how to create a cognitive user model to estimate latent cognitive variables such as cognitive load or free mental working memory capacity. We have presented a concept for a cognitive user model which is based on Hierarchical Bayesian Models (HBM). The model has been validated numerically using synthetic data. The final model is able to explain the three observable variables task success, mission score and mission time with an accuracy of over 90%. The generic structure of the modeling approach allows transferability to other application fields.

Several issues with the final model have been identified and provide directions for further research. The presented results are based on simulated data, not on empirical data, hence, an evaluation by a user study is the next logical step. Further work could also incorporate more observable variables, e.g., data from eye-tracking or physiological sensors.

References

1. Advanced Distributed Learning (ADL): Experience API (xAPI) Specification, Version 1.0.1. Tech. rep., Advanced Distributed Learning (ADL) Initiative, U.S. Department of Defense (2013)
2. Aydinbas, M.: Realizing Cognitive User Models for Adaptive Serious Games. Master thesis, TU Darmstadt, Fraunhofer IOSB (2019)
3. Conati, C., Lallé, S., Rahman, M.A., Toker, D.: Comparing and combining interaction data and eye-tracking data for the real-time prediction of user cognitive abilities in visualization tasks. ACM Trans. Interact. Intell. Syst. **10**(2) (2020)
4. Conati, C., Maclaren, H.: Empirically building and evaluating a probabilistic model of user affect. User Modeling & User-Adapted Interaction **19**(3), 267–303 (2009)

5. Dörner, R., Göbel, S., Effelsberg, W., Wiemeyer, J.: Serious Games - Foundations, Concepts and Practice. Springer International Publishing, Cham (2016)
6. Farrell, S., Lewandowsky, S.: Computational Modeling of Cognition and Behavior. Cambridge University Press (2018)
7. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian Data Analysis. Chapman and Hall/CRC, third edn. (2014)
8. Hallifax, S., Lavoué, E., Serna, A.: To tailor or not to tailor gamification? an analysis of the impact of tailored game elements on learners' behaviours and motivation. In: Artificial Intelligence in Education. pp. 216–227. Springer Int. Pub. (2020)
9. Kickmeier-Rust, M., Albert, D.: Educationally adaptive: Balancing serious games. International Journal of Computer Science in Sport **10**(1), 1–10 (2012)
10. Laird, J.E.: The SOAR cognitive architecture. AISB Quarterly **134**(1), 1–4 (2012)
11. Lambert, B.: A Student's Guide to Bayesian Statistics. Sage Pub. Ltd (2018)
12. van Merriënboer, J.J.G., Sweller, J.: Cognitive load theory and complex learning: Recent developments and future directions. Educational Psychology Review **17**(2), 147–177 (2005)
13. Pelánek, R.: Metrics for evaluation of student models. International Educational Data Mining Society **7**, 1–19 (2015)
14. Seyderhelm, A.J.A., Blackmore, K.L., Nesbitt, K.: Towards cognitive adaptive serious games: A conceptual framework. In: van der Spek, E., Göbel, S., Do, E.Y.L., Clua, E., Baalsrud Hauge, J. (eds.) Entertainment Computing & Serious Games. pp. 331–338. Lecture Notes in Computer Science, Springer Int. Publishing (2019)
15. Shute, V., Zapata-Rivera, D.: Adaptive educational systems. Adaptive technologies for training and education **7**(1), 1–35 (2012)
16. Sweller, J.: Element interactivity and intrinsic, extraneous, and germane cognitive load. Educational Psychology Review **22**(2), 123–138 (Jun 2010)
17. Sweller, J.: Cognitive load theory. Psychology of Learning & Motivation **55**, 37–76 (2011)
18. Tadlaoui, M.A., Carvalho, R.N., Khaldi, M.: A learner model based on multi-entity bayesian networks and artificial intelligence in adaptive hypermedia educational systems. Int. Journal of Advanced Computer Research **8**(37), 148–160 (2018)
19. Woolf, B.P.: Building Intelligent Interactive Tutors. Morgan Kaufmann (2009)
20. Wray, R.E., Woods, A.: A cognitive systems approach to tailoring learner practice. Conf. on Advances in Cognitive Systems ACS **21**, 18 (2013)