Kickstarting Adaptive Fact Learning Using Hierarchical Bayesian Modelling

Maarten van der Velde[1] **(m.a.van.der.velde@rug.nl)**
Florian Sense[1] **(f.sense@rug.nl)**
Jelmer Borst[2] **(j.p.borst@rug.nl)**
Hedderik van Rijn[1] **(d.h.van.rijn@rug.nl)**
[1]Dept. of Experimental Psychology & Behavioural and Cognitive Neuroscience, University of Groningen, The Netherlands
[2]Bernoulli Institute, dept. of Artificial Intelligence, University of Groningen, The Netherlands

Learning facts is an inescapable part of education, whether it be memorising French words or studying US topography. Our lab has developed a digital learning environment that uses a cognitive model of human memory to determine when and how often a fact should be rehearsed during a learning session. The system tracks how difficult each fact is for a given student, continually refines this assessment on the basis of the student's responses, and adjusts the scheduling of items so that difficult facts are repeated sooner and more frequently than easy facts. This adaptive fact learning system has been successfully applied in various contexts (van Rijn, van Maanen, & van Woudenberg, 2009; Sense, van der Velde, & van Rijn, 2018).

Currently, all facts are initially assumed to be equally difficult for all learners. As observations are made, the difficulty estimate is tuned to the right level for each learner-fact pair. This means that knowledge about a learner's general ability or about a fact's typical difficulty is not used at all. In this study, we propose several methods for using data from prior learning sessions to inform the initial difficulty estimates of the model. Using such learning history is expected to make the learning process more efficient, as the model would be better able to quickly hone in on the appropriate difficulty estimate for each fact. We use hierarchical Bayesian modelling to make individualised predictions on the basis of previous learning sessions and test these predictions in a new session.

Adaptive fact learning model

The scheduling of items within a learning session is determined by an adaptive model that builds on earlier work by Pavlik and Anderson (2005). This model is described in more detail in Sense, Behrens, Meijer, and van Rijn (2016).

The model represents each fact by its own memory chunk, with an activation (a measure of the strength of the memory trace) that is boosted by each repetition and decays over time. At time t, and given n previous repetitions at $t_1,\ldots,t_n$ seconds ago, the activation A of chunk i is expressed by Equation 1. The d parameter in this equation controls how quickly a fact's activation decays after a repetition, and therefore how frequently the fact is repeated. Differences in difficulty between facts are captured in the *rate of forgetting* parameter α, a component of d, which is estimated separately for each learner-fact pair. The more difficult a fact is, the higher its rate of forgetting will be, and the faster its activation will decay.

$$A_i(t) = ln\Big(\sum_{j=1}^{n} t_j^{-d_i(t)}\Big) \quad \text{with} \quad d_i(t) = 0.25 * e^{A_i(t_{n-1})} + \alpha_i \qquad (1)$$

At any given time, the system selects whichever fact has the lowest estimated activation to be rehearsed, thereby maximising the spacing between repetitions while also aiming to repeat each fact before it is forgotten. A new fact is introduced only when all activation values are above a threshold of -0.8.

The system currently starts out with the assumption that all facts have a rate of forgetting of 0.3, and it refines this estimate over the course of the learning session. It uses the difference between expected response times (based on the fact's activation at the time of presentation; the higher the activation, the faster the expected response) and observed response times, as well as response accuracy, to make step-wise adjustments to the estimate that best reflect the observed behaviour.

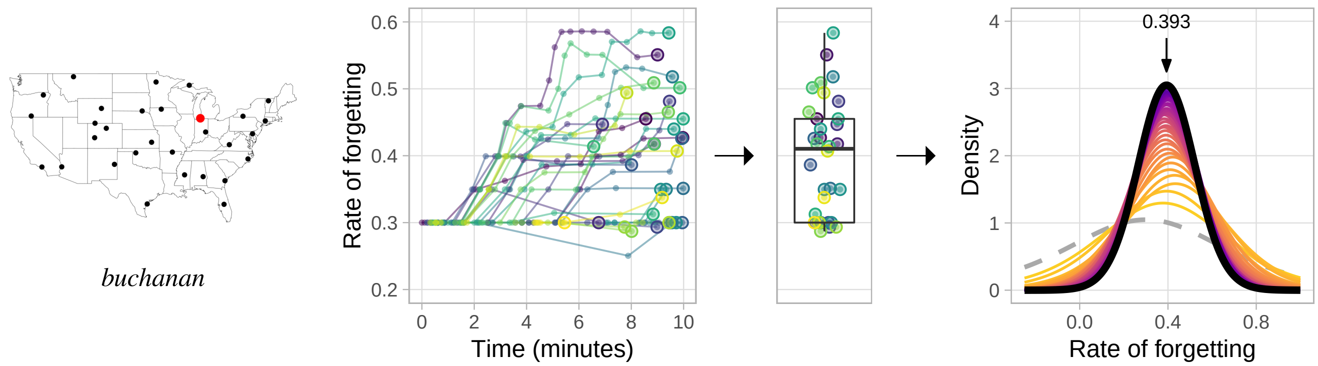
Predicting rate of forgetting

In this study, we use previous learning history to predict what the rate of forgetting of a particular fact will be for a given learner. We then take this prediction as the initial rate of forgetting estimate, rather than the default value.

We test four prediction methods and compare them to the default prediction of 0.3. Fact-level difficulty estimates for a set of topography facts (names of relatively unknown US cities; see Figure 1a for an example) were obtained from an initial experiment in which participants completed a learning session with the default system. In a follow-up experiment, learner-level estimates were derived for different participants who studied a comparable set of facts with the default system. These participants then completed another learning session in which they studied the facts from the first experiment with a system that, depending on the condition to which they were assigned, initialised new facts with a rate of forgetting that was based on one of the four prediction methods or on the default value[1].

Fact-level prediction As multiple learners study the same fact, we form an increasingly detailed picture of its difficulty through the rates of forgetting observed in all these learners. It is to be expected that a new learner studying this fact will

[1] A preregistration with a more detailed description of the protocol and the analysis plan is available at https://osf.io/vwg6u/.



(a) (b) (c) (d)

Figure 1: The process by which a fact-level prediction of the *rate of forgetting* of the fact shown in **(a)** is made. **(b)** In previous learning sessions, the rate of forgetting of the fact is estimated separately for each learner and refined over the course of the session. **(c)** The final estimates (one per learner) are used to train the Bayesian model. **(d)** The posterior predictive distribution of the Bayesian model is updated as observations are added. The prior predictive distribution is shown as a dashed grey line, the final posterior predictive distribution as a black line, with intermediate predictive distributions shown in increasingly dark colours. The model prediction, indicated by the arrow, is the mode of this final distribution.

find it similarly difficult. For this reason, we use the rates of forgetting measured in other learners to make a fact-level prediction of the rate of forgetting that can be used as an initial estimate when the fact is encountered by a new learner. Figure 1 shows how such a prediction is made. Predicted rates of forgetting come from a hierarchical Bayesian model which models the rate of forgetting using a Normal-Gamma distribution with a weakly informative prior centered on 0.3: $NG(\mu = 0.3, \kappa = 3, a = 1, b = 0.2)$.

Learner-level prediction Through the same process as in the fact-level prediction, but instead using the rates of forgetting of all facts that a given learner has encountered in the past, we can predict a learner's rate of forgetting. This value is then used as the initial estimate for all facts that the learner encounters.

Fact- and learner-level prediction We also test a method in which a distinct prediction is made for each learner-fact pair. Two posterior predictive distributions—one for the fact-level prediction and another for the learner-level prediction—are combined using logarithmic opinion pooling (Genest, Weerahandi, & Zidek, 1984) with equal weights. The mode of this combined distribution becomes the predicted rate of forgetting.

Domain-level prediction The domain-level prediction, reflecting the general difficulty of the material in a domain among a certain population, is the mean of all fact-level predictions for the set of facts. This value is used as the initial rate of forgetting in all learner-fact pairs, resulting in a domain-specific alternative to the fixed default prediction of 0.3.

Results & Discussion

Data have been collected from 159 participants for the second experiment, which tests the predictions made by the Bayesian model, while a replication in an online sample is still ongoing. Preliminary results suggest that using learning history to predict rates of forgetting does affect learning performance, as participants are more accurate while studying if the system uses one of the prediction methods (a Bayesian ANOVA shows strong evidence for an effect of condition on accuracy: $BF_{10} = 15.7$), potentially with beneficial effects on motivation. However, this does not appear to translate to higher performance on a delayed recall test (a Bayesian ANOVA shows strong evidence against an effect of condition on test score: $BF_{01} = 19.9$). We will conduct further analyses to address the other questions set out in the preregistration[1], as well as any exploratory questions that arise from this rich data set.

References

Genest, C., Weerahandi, S., & Zidek, J. V. (1984). Aggregating opinions through logarithmic pooling. *Theory and Decision*, *17*(1), 61-70.

Pavlik, P. I., & Anderson, J. R. (2005). Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect. *Cognitive Science*, *29*(4), 559-586.

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An Individual's Rate of Forgetting Is Stable Over Time but Differs Across Materials. *Topics in Cognitive Science*, *8*(1), 305-321.

Sense, F., van der Velde, M., & van Rijn, H. (2018). Deploying a Model-based Adaptive Fact-Learning System in University Courses. In *Proceedings of the 16th International Conference on Cognitive Modeling* (p. 136-137).

van Rijn, H., van Maanen, L., & van Woudenberg, M. (2009). Passing the test: Improving learning gains by balancing spacing and testing effects. In *Proceedings of the 9th International Conference on Cognitive Modeling* (p. 110-115).