Optimizing Multi-Context Motor Adaptation using Model-Free Reinforcement Learning

The efficacy of motor rehabilitation is often limited by poor generalization from training tasks to real-world movements. To understand this phenomenon one needs to understand the generalization in motor learning better overall: how learning one movement affects similar movements[1]. In motor learning research one often works with experiments consisting of several movement goals=sub-tasks (e.g. baseline, perturbation, washout, different targets). Some studies focused on studying so-called "structural learning"[2]: testing which sequence of various sub-tasks (we call them "contexts") presentations allows for the best learning performance. To our knowledge, such sequences have been always chosen heuristically (e.g. various alternations of blocks of fixed size of different preselected deterministic perturbations or of random perturbations). In this research we look for training schedules that are principled, not heuristic. Can we computationally discover an optimal training curriculum that accelerates the acquisition of consistent motor performance?

We introduce a novel framework where a model-free deep reinforcement learning (RL) agent learns to sequence movement contexts for a simulated human learner. Crucially, we redefine the learning objective away from minimizing average error. Instead, the agent is trained to minimize a lower-quantile of the error distribution (e.g., the 40th percentile). This objective prioritizes the achievement of reliable, consistent performance, which is more representative of functional skill. Our "virtual subject" is an advanced state-space model capturing important features of motor adaptation, including learning rate adaptation and context-dependent generalization.

The RL-tutor discovers non-intuitive policies that significantly outperform both random and heuristic "smart" schedulers, especially in high-interference scenarios where similar movements require conflicting adaptations. By optimizing for performance consistency, our framework learns curricula that better manage motor interference. This work presents a principled, automated method for designing training protocols that exploit the structure of motor learning to achieve robust outcomes, offering a promising new direction for personalized rehabilitation.

Additional Detail

The central question is how to sequence a curriculum of motor tasks to optimally accelerate the acquisition of reliable motor skill. We formalize this as a policy discovery problem and solve it using a model-free reinforcement learning (RL) agent that tutors a computational model of a human learner.

- **Virtual Subject Model:** To simulate a human learner, we use a state-space model that is a mixture of two previously existing models [1,3]. The model's internal state x is in context i is updated based on performance error e in context j on trial t: $x^i_{t+1} = x^i_t + S^{ij}\lambda_t\left(e^j_t\right)e^j_t$, where S^{ij} is the similarity score (between 0 and 1) of contexts i and j, $\lambda_t(e)$ is the error sensitivity (adaptive learning rate) evaluated after trial t as in Memory of Errors model[3], meaning that it increases when two preceding trials' errors had the same sign, and decreases otherwise.

- **RL-Tutor Agent:** We frame the context scheduling environment as a Markov Decision Process with a model-free RL agent scheduling contexts for the virtual subject. The reinforcement learning agent is trained using the REINFORCE algorithm [4] with a reward consisting of the scaled 40th percentile error awarded to the model after each trial.

We evaluated three schedulers in a simulated multi-target reaching task: (1) a naive random policy, (2) a Similarity-Conflict-Adjusted (SCA) heuristic policy, and (3) the trained RL-tutor. The SCA policy is a Markov chain that favors exposure to similar contexts with low conflict and large perturbations. Our results demonstrate the superiority of optimizing for a quantile-based metric. As shown in Figure 1 left column, the RL-tutor discovers a policy that more rapidly reduces the 40th percentile error compared to the other methods, particularly in high-interference environments. While the naive random scheduler is sampling similar, high-conflict context pairs (Fig. 1 right column), the RL-tutor learns to intelligently interleave these with other contexts to manage interference and foster robust learning. This leads to a marked improvement in the consistency of performance across the subset of contexts.

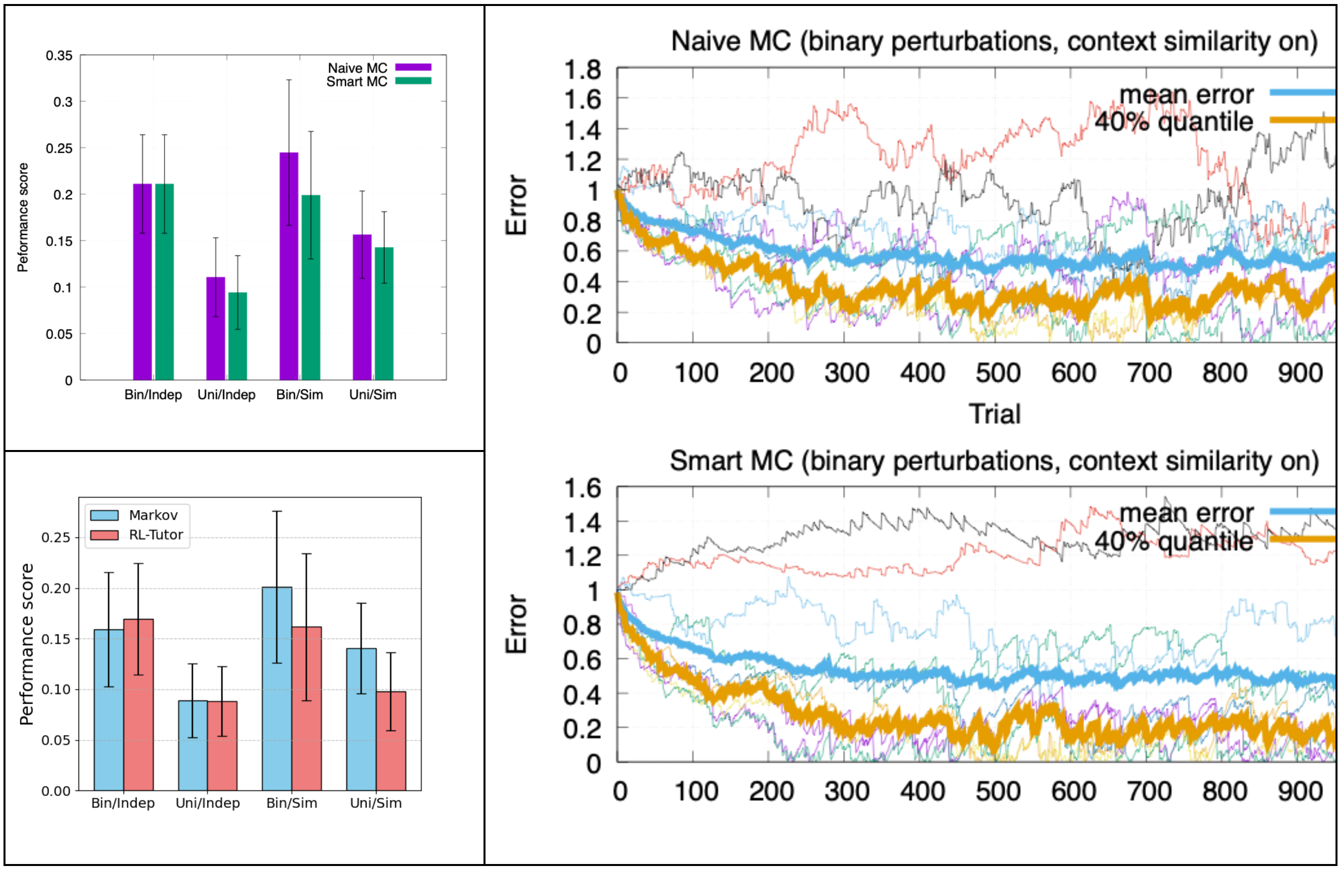

Figure 1. Left: comparison between three sequencing paradigms: equiprobable contexts (Naive MC, Markov), reduced probability of contexts with small perturbations/high conflict scores (Smart MC), and RL-based. Bin: binary perturbations (+/- 1); Uni: uniformly distributed perturbations [-1:1]; Indep: independent contexts (zero similarity: $s^{ij} = 0$) ; Sim: locally similar contexts (adjacent contexts' similarity = 1/2). **Right: dynamics of individual errors in each context, mean absolute error (blue) and the 40th percentile absolute error (orange) in case of binary perturbations (+/- 1) and locally similar contexts for the random (top) and SCA (bottom) sequencers.**

Significance and Implications: This work provides a proof-of-concept for automated discovery of optimal training curricula in motor learning. It contributes on one hand, to motor learning research by generating novel, testable predictions (e.g. that specific non-intuitive training sequences discovered by the RL outperform conventional schedules). And this research also contributes to computational neuroscience & NeuroAI by using together a descriptive model of neural adaptation with a normative, model-free control algorithm. We believe that this research is a stepping stone for designing better rehabilitation protocols

References:

1. Donchin O, Francis JT, Shadmehr R. Quantifying Generalization from Trial-by-Trial Behavior of Adaptive Systems that Learn with Basis Functions: Theory and Experiments in Human Motor Control. J Neurosci. 2003;23:9032–45.
2. Braun DA, Aertsen A, Wolpert DM, Mehring C. Motor Task Variation Induces Structural Learning. Curr Biol. 2009;19:352–7.
3. Herzfeld DJ, Vaswani PA, Marko MK, Shadmehr R. A memory of errors in sensorimotor learning. Science. 2014;345:1349–53.
4. Williams RJ. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach Learn. 1992;8:229–56.