Using Behavioural Cloning and RL to Teach Paired Associates

Abstract

We study the problem of curriculum design for associative learning. We frame curriculum design as a sequential decision problem: a teacher chooses which paired associates to present to a learner next, conditioned on the learning history. A good curriculum should maximize the learner's retention at the end of learning. Our pipeline has five stages: (1) collect human learning trajectories under a diverse set of curricula, (2) train predictive models to mimic human learning trajectories ("behavioural clones"), (3) use reinforcement-learning to compute optimal curricula for these learner models, (4) interpret learned policies, and (5) evaluate on new human cohorts. We report preliminary results using a policy gradient approach to learn a curriculum for teaching simple cognitive models of associative learning (rather than full behavioural clones). The learned curriculum outperforms baselines, including random scheduling and Leitner spaced repetition. Unlike existing spaced repetition schedules, whose policies are not typically conditioned on time, our learned policy is time-dependent, first focusing on consolidating a subset of associations early in training, and then quickly "cramming" the remaining cues late in training. These results provide a proof of concept that deep RL can be used for data-driven curriculum design, discovering adaptive sequencing strategies beyond classic heuristics.

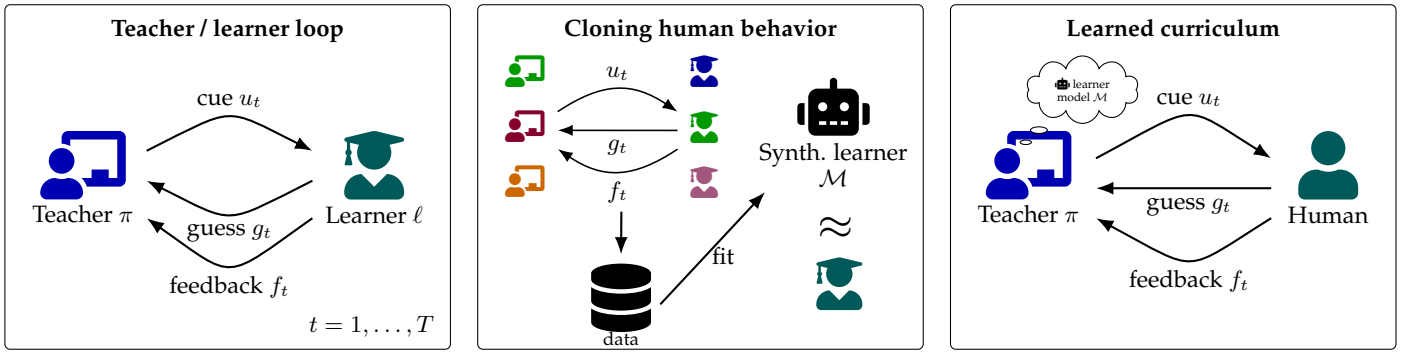

Figure 1: Overview of the paired-associates teaching setup and pipeline. **left:** teacher–learner loop over learning trials $t = 1, \ldots, T$, where the teacher policy π selects a cue u_t, the learner ℓ produces a guess g_t, and the teacher provides feedback $f_t = \psi(u_t)$. **centre:** behavioural cloning: trajectories collected from multiple curricula and learners are used to fit a synthetic learner model $\mathcal{M}$. **right:** A teacher leverages a learner model $\mathcal{M}$ to optimally teach a human or artificial learner.