**MACHINE TEACHING OF ACTIVE SEQUENTIAL LEARNERS**

**Tomi Peltola, Mustafa Mert Çelikok, Pedram Daee, Samuel Kaski**
Helsinki Institute for Information Technology HIIT, Department of Computer Science, Aalto University, Finland

**INTRODUCTION**

Machine teaching: Find the best training data that can guide a learning algorithm to a target model with minimal effort.
- Traditionally, the teacher provides data by sampling labels from the true data distribution (consistent teacher).
- Providing true labels can be sub-optimal in finite-horizon tasks for sequential learners that actively choose their queries.

**Mean Cumulative Reward**

**Model of the Learner**
- Random draw (baseline)
- Human Teacher - Naive Model
- Planning (P-P; left-side panels)

We address the complementary problem of teaching-aware learning by endowing the learner with a model of the teacher.
- The performance increases markedly when the learner models the teacher.
- The planning teacher (P; Equation 4 or Equation 3) based on the ground truth likelihood. Learner thinks the teacher is:
- Naive
- Planning

Planning

We formulate this sequential teaching problem, as an MDP, and allow the teacher to provide data inconsistent with the true distribution ("With teacher" panel on the right).
- We address the complementary problem of teaching-aware learning by endowing the learner with a model of the teacher.
- The final inference problem reduces to inverse reinforcement learning.

**Teacher Models**
- Naive:
  - $p_{\text{M}}(a_t | i, \pi) = \text{Bernoulli}(a_t | \pi_i)$
- Planning:
  - $p_{\text{M}}(a_t | i, M_t, w) = \exp \left( \sum_{k} \exp \left( R_{M_k}(a_t, w') \right) \right)$
- Mixture:
  - $p_{\text{M}}(a_t | i, M_t, w, \pi, \alpha) = \alpha p_{\text{M}}(a_t | i, M_t, w) + (1-\alpha) p_{\text{M}}(a_t | i, \pi_i)$

**EXPERIMENTS**

Setup:
- Word search study: the teacher selects a target word and the learner tries to guess the word by asking sequential questions.
- Learner: "Is this word relevant to the target?", Teacher: Yes/No

Results with Human Teachers:
- Participants ($\alpha = 1$) achieved noticeably higher rewards while interacting with a learner having the mixture teacher model (red), compared to the naive teacher model (blue).

Results with Simulated Teachers:
- The planning teacher can steer a teacher-unaware learner to achieve a marked increase in performance compared to a naive teacher (P-N vs N-N; left-side panels)
- The performance increases markedly when the learner models the planning teacher (P-P; left-side panels)

**CONCLUSION**

- We have introduced a new sequential machine teaching problem, where the learner actively chooses queries (e.g., in active learners and multi-armed bandits) and the teacher provides responses. The new teaching problem is formulated as a Markov decision process, where the solution provides the optimal teaching policy. Using the MDP formulation, teacher-aware learning from the teacher’s responses is formulated as probabilistic inverse reinforcement learning.
- The proposed teaching framework holds promise for a feasible and natural computational approach in modelling active user behaviour in interactive intelligent systems.

See the paper website for more info and the code: [https://git.io/JeSaU](https://git.io/JeSaU)