Some Title

Authors

July 22, 2025

1 Introduction

>>> To be completed...

2 Related Literature

>>> To be completed...

3 Experimental Setup

Two experimental procedures, one aiming to assess the efficacy and scalability of the coaching process and one to explore a more realistic and open–ended environment.

Add a few words about how coaching happens in the context of searching, i.e., how from a past sub–optimal state the coach provides a piece of advice according to their (implicit / explicit) policy that improves the state towards a(n implicit / explicit) goal.

3.1 Sorting Coaching

In this series of coaching sessions we have implemented two variants of a machine coach, aiming to explain to an initial ignorant agent about how to sort a list of numbers, as a rough equivalent to linear preference elicitation. The first coach, $\mathcal{C}_{\text{bubble}}$, uses bubble sort as its underlying policy while the second one, $\mathcal{C}_{\text{quick}}$, uses quicksort to provide advice from. We explored different state sizes, n, ranging from 1 to 20, running $m = 100$ iterations for each value of n. We have also explored two different types of advice for each coach: (A1) *full advice*, where each piece of advice depends on the entire state, and; (A2) *partial advice*, where each piece of advice depends on the specific part of the state to be improved. Moreover, for each advice type we have considered three learner configurations regarding advice memory: (C1) *no memory* across different iterations for the same state size, n, i.e., coaching sessions are pairwise independent; (C2) *short memory* across different iterations for the same state size, n, i.e., coaching sessions across different values of n are independent but not within the same value for n, and; (C3) *long memory* across all values of state size, n, so all coaching sessions are correlated, building on knowledge from previous test cases.

4 Results

>>> To be completed...

5 Conclusions and Future Work

>>> To be completed...