

**Отзыв научного руководителя  
о выпускной квалификационной работе  
Тыщука Кирилла Ильича**

**«Обнаружение скрытых шаблонов в обучении с подкреплением»**

Работа Кирилла Тыщука посвящена крайне актуальной в настоящее время области: обучению с подкреплением, в частности выделению паттернов из учебных демонстраций эксперта. Обучение с подкреплением — это раздел машинного обучения, в котором обучающийся агент старается максимизировать ту или иную целевую функцию (награду) без готового обучающего набора данных, а в интерактивном взаимодействии с окружающей средой. Например, именно так обучаются многие модели машинного обучения в робототехнике: нет датасета, который бы содержал реакции среды на действия робота (выраженные, например, в импульсах двигателей), но есть возможность пробовать совершать эти действия в симуляционной среде.

Из этой постановки задачи следует актуальность и той конкретной задачи, которая рассмотрена в диссертации: имитационного обучения на основе малого количества экспертных данных. К примеру, мы не можем создать для проходящего лабиринт агента датасет из миллиона последовательностей, где человек-эксперт проходит разнообразные лабиринты (это было бы слишком дорого), но можем породить несколько таких примеров, которые должны стать для робота некоторым образцом. Более того, вполне естественно, что во многих ситуациях мы обладаем лишь частичной информацией об этих экспертных последовательностях состояний: например, у нас есть ролики с записью прохождения лабиринта, но нет информации о том, какие кнопки в какой момент эксперт нажимал. См., например, работу Aytar et al. (2018)<sup>1</sup>.

В работе Кирилла задача конкретизируется следующим образом: на основе идеи иерархического обучения с подкреплением, где решение задачи разбивается на «крупные» стратегические и «мелкие» конкретные действия, требуется извлечь из экспертных траекторий те самые «крупные действия», паттерны, которые могут дать информацию об окружающей среде и помочь выработать стратегию агента. Кирилл придумал и реализовал новое окружение в формате OpenAI Gym (коллекция сред для обучения с подкреплением), в котором требуется посетить заданные клетки поля («кнопки») в порядке, соответствующем их цветам. Таким образом, естественным понятием «крупного действия» здесь является переход от одной кнопки к другой, т.е. нужная последовательность цветов и местоположение соответствующих кнопок.

Основной технической идеей работы является переход от обучения непосредственно стра-

---

<sup>1</sup>Aytar, Yu. et al. Playing hard exploration games by watching YouTube. NeurIPS 2018.

тегии (функции, показывающей, какое действие нужно совершить в данном состоянии) к обучению ранжирующей модели  $f(s, s')$ , показывающей, насколько выгодным ожидается переход из состояния  $s$  в  $s'$ . Идея здесь в том, что такая функция может показывать в том числе и более далёкие цели агента, недостижимые за один переход, и может быть использована для выделения «крупных действий».

Ранжирующая модель в работе строится на основе сиамских сетей, точнее, классической модели DSSM<sup>2</sup>, часто используемой в информационном поиске. Для этой модели предложен вариант функции потери FPS loss на основе метода contrastive predictive coding<sup>3</sup>, а также добавлена квантизация в конце одной из веток DSSM (это, насколько мне известно, новая идея, ранее не встречавшаяся в литературе). Кроме того, для интерпретации полученных результатов разработан специальный метод визуализации на основе тепловых карт (heatmaps).

В результате на основе широкого и содержательного эмпирического исследования удалось показать, что модель с квантизацией действительно обучается более полезным высокоуровневым концептам, и в выбранной среде обучения показывает отличные результаты в сеттинге имитационного обучения. Хочется особо отметить самостоятельность и активность Кирилла в процессе работы над диссертацией: вся реализация и подавляющее большинство идей, приведённых в работе, составляют исключительно его личную заслугу, и в этой работе Кирилл показал себя как самостоятельный и высоко мотивированный исследователь.

Таким образом, по моему мнению, выпускная квалификационная работа К. И. Тыщука удовлетворяет всем требованиям и представляет собой законченное исследование высокого профессионального уровня. Считаю, что Тыщук Кирилл Ильич заслуживает присуждения ему степени бакалавра, а работа заслуживает оценки «отлично».

Научный сотрудник Санкт-Петербургского  
отделения Математического института  
им. В.А. Стеклова Российской академии наук,  
кандидат физико-математических наук

С. И. Николенко

---

<sup>2</sup>Huang et al. Learning deep structured semantic models for web search using clickthrough data. CIKM 2013.

<sup>3</sup>Van den Oord et al. Representation learning with contrastive predictive coding. arxiv:1807.03748.