**EAWC Assignment 2**

Scientific paper analyzing report

**Student ID: 2023312829**

**Name: Ge Haizhou**

1. **Basic information**

Title: Eureka: Human-Level Reward Design via Coding Large Language Models

Author: Yecheng Jason Ma, William Liang, Guanzhi Wang, De-An Huang, Osbert Bastani, Dinesh Jayaraman, Yuke Zhu, Linxi Fan, Anima Anandkumar

Source: https://doi.org/10.48550/arXiv.2310.12931

1. **Special terms**

* **Large Language Models:** A large language model (LLM) is a large-scale language model notable for its ability to achieve general-purpose language understanding and generation.
* **Manipulation**: Manipulation refers to an agent's control of its environment through selective contact.
* **Reinforcement learning (RL)**: RL is an interdisciplinary area of machine learning and optimal control concerned with how an intelligent agent ought to take actions in a dynamic environment to maximize the cumulative reward.
* **Shadow Hand:** The Shadow (Dexterous) Hand is a humanoid robot hand system developed by The Shadow Robot Company in London. The hand is comparable to a human hand in size and shape and reproduces all of its degrees of freedom.

1. **Move-step approach and key sentence patterns in each section**
   1. **Abstract**

**3.1.1 Move-step approach**

Sentence (1) indicates Move 1 by showing the status of LLM. Sentence (2) indicates Move2 and Move3 by showing learning complex low-level manipulation tasks remains an open problem. Sentence (3) indicates Move4 by presenting the EUREKA algorithm. Sentences (4-5) indicate Move5 by showing it exploits LLMs to write reward code. Sentences (6-9) indicate Move6 by showing its better performance and contributions. Sentence (6) indicates Step1 by simply describing the excellent performance. Sentence (7) indicates Step2 by showing the improvement compared with human experts. Sentence (8-9) indicate Step3 by showing the algorithm can enable RLHF and can be applied in performing pen spinning tricks with a simulated Shadow Hand.

**3.1.2 Key sentence patterns**

* However, harnessing … to …, such as …, remains an open problem.
* We bridge this fundamental gap and present …, a...
  1. **Introduction**

**3.2.1 Move-step approach**

Sentence (1) indicates Move1 and Move2 by first showing the territory of LLM and then raising the question whether they can be used to learn complex low-level manipulation tasks, which is Step 1C. Sentence (2) indicates Move2 Step 1E by pointing out the limitations of existing attempts. Sentence (3) indicates Move1 by generalizing about reinforcement learning. Sentence (4-5) indicates Move2 Step 1B by showing the difficulty of designing rewards. Sentence (6-8) indicates Move2 Step 1C by asking whether it is possible to develop a universal reward programming algorithm using state-of-the-art coding LLMs. Sentences (9-end) indicate Move3 by introducing the EUREKA algorithm. Sentence (9) describes the present research that is powered by coding LLMs. Sentences (10-11) correspond to Step 5 by showing the results compared with experts. Sentences (12-17) and the last paragraph state the contributions in several aspects, which correspond to Step6.

**3.2.2 Key sentence patterns**

* Existing attempts require…, leaving a substantial gap in achieving…
* On the other hand, … has achieved impressive results in …
* Despite their fundamental importance, …are known to be notoriously difficult to …
* Given the paramount importance of …, we ask whether it is possible to develop …
  1. **Methods**

**3.3.1 Move-step approach**

Sentence (1) indicates Move1 Step5 by describing the three components of the EUREKA model. Sentences (2-14) indicate Move1 Step7 by justify the procedures by two reasons and highlighting the zero-shot advantages. Sentences (15-16) indicate Move2 Step1 by showing how the input data are sampled. Sentences (17-18) explain the bugs exponentially decreases as the number of samples increases, indicating Move2 Step2. Sentences (19-25) indicate Move2 Step3 by demonstrating how evolutionary search presents a natural solution that addresses the execution error and sub-optimality challenges. 6 3 Sentences (26-34) indicate Move3 by describing the procedure of reward reflection. Sentences (26-31) recount the importance of putting into words the quality of the generated rewards which indicate the Step1, while sentences (31-34) showing that by providing detailed accounts on how well the RL algorithm optimizes individual reward components, reward reflection enables EUREKA to produce more targeted reward editing and synthesize reward functions that better synergize with the fixed RL algorithm.

**3.3.2 Key sentence patterns**

* … consists of three algorithmic components: …
* That is, …
* This is intuitive for two reasons: First, …
* Remarkably, with only these minimal instructions, …
  1. **Results**

**3.4.1 Move-step approach**

Sentences (1-2) shows the aggregate results are compared on two benchmarks, indicating Move1. Sentences (3-7) indicate Move2 by comparing the results from EUREKA and other models. Sentence (8) indicate Move3 by showing that the outcomes indicate that EUREKA’s general principles can be readily applied to coding LLMs of varying qualities. Sentence (9) indicates Move1 by showing the main purpose of these results is demonstrating EUREKA consistently improves over time. Sentences (10-14) indicates Move2 by describing the average performance of the cumulative best EUREKA rewards after each evolution iteration. Sentence (15) indicates Move3 by showing these results demonstrate that EUREKA’s novel evolutionary optimization is indispensable for its final performance.

**3.4.2 Key sentence patterns**

* In Figure …, we report the aggregate results on the … benchmarks.
* Despite being provided access to some of, … still underperforms …
* We hypothesize that …, leaving more room for …
* In a few cases, … but perform significantly better, demonstrating that …
  1. **Discussion**

There is no discussion in this paper.

* 1. **Conclusion**

**3.4.1 Move-step approach**

Sentence (1) indicates Move1 by telling the presentation of EUREKA. Sentences (2-4) indicate Move2 by summarizing the advantages of this approach. Sentence (5) suggests that the simple principle of combining large language models with evolutionary algorithms is a general and scalable approach to reward design, an insight that may be generally applicable to difficult, open-ended search problems.

**3.4.2 Key sentence patterns**

* … achieves … on a wide range of robots and tasks.
* …, readily incorporating … and … to …

1. **My review**

Overall, the paper presents a novel approach to reward design for reinforcement learning using LLMs, demonstrating impressive results in various environments. The generality of EUREKA and its ability to incorporate human feedback make it a promising advancement in the field of robotic manipulation and reinforcement learning. However, a detailed analysis of potential limitations, scalability, and real-world applicability would further strengthen the paper.