Homework 5: Paper 1 Review

Hyuntaek Oh

Due Nov 26, 2024

1 SUMMARY

1.1 What is the paper about?

In this paper, the authors propose an algorithm called Reactive Multi Fitness Learning (R-MFL) for solving the problem of longitudinal tasks (sparse rewards) and providing stable reaction to unforeseen changes with non-forgetting. Then, they evaluate the R-MFL algorithm by comparing with conventional MFL learning algorithm.

1.2 Main contributions

The main contributions of this paper are to introduce a policy structure that leverages multiple agent behaviors to learn a long-term sparse reward by:
1) injecting new behaviors into the system to react to unforeseen disturbances
2) analyzing multiple local rewards to identify when and where an agent needs a reactive injection of new behaviors.

1.3 The paper's strengths

The paper's strengths are to start with the story that reaches why their approach is meaningful and needed and to focus on how to enable multi-robot teams to adapt to dynamic environments and sparse reward structures during long-term tasks without catastrophic forgetting. The paper identifies key gaps in existing methods such as non-adaptation skills in Multi Fitenss Learning algorithm and provides a novel solution called Reactive Multi-Fitness Learning (R-MFL). The proposed R-MFL framework is clearly described, with detailed pseudo codes for key steps. The simulation data is reliable by showing the comparison and the performance of two different algorithms.

1.4 The paper's weaknesses

Mostly, the paper is well-defined and documented. Nevertheless, the paper mentioned some limitations on their algorithms but does not discuss them in depth. The scalability to larger teams and more complex environments is not explicitly addressed, leaving questions about its generality.

1.5 What are the conclusions?

The authors conclude that R-MFL algorithm can provide agents with robust, stable reaction to unforeseen circumstances by showing experiment results implying the benefits of using it. There are benefits of using R-MFL: identifying which behavior populations need a new behavior based on locally available rewards, requesting by an agent to a new policy to augment its capabilities and overcome environmental disturbances, utilizing atomic behaviors and value iteration on population of similar behaviors in adaptation, coordinating agents behavior without forgetting and deteriorating overall performance, and helping the adaptation process by singling out which agent and which behaviors need adapting.

2 GENERAL RATING: Good

2.1 Presentation, Readability, and Organization

This paper is well-organized and follows the order of categories in usual science research paper, including abstract, introduction, related works, method, experiment, results, discussion, conclusion, and references. The authors told story of why the R-MFL algorithm is necessary to solve the longitudinal, sparse reward problems, improving the limitations of previous works. The paper is easy to read even though there are some words hard to understand without background knowledge such as the meaning of catastrophic forgetting. There are only a few minor grammatical errors and typos. In Reactive multi-fitness learning section, 'D. step 3: Reacting With New Behaviors' should be 'D. step 4: ... '.

2.2 Problem definition

The authors described big problems that multi-robot systems have been used for remote exploration and multiple tasks. In particular, longitudinal tasks is problematic since the robots need to handle both environmental and their behavioral changes in the system. The main problem it provides is that the delay between actions and rewards become longer, making hard to determine their action selection. This paper offered practical algorithms to address independent training of behavior population, training of R-MFL top level policies using CCEA Neuro-Evolution, identifying when and how to react, and R-MFL during deployment with value iteration. The explanation about these algorithms was clear and detailed to make readers understand the meaning of each step at certain lines.

2.3 Originality

The proposed work is original since previous work like MFL was able to learn and operate in the sparsely rewarded multiagent settings this paper looks at. However, fundamentally it has no method to react to changed while agents are deployed. Furthermore, MFL uses a two-step learning algorithm whereas R-MFL has three main steps to generate pre-training behaviors, train top-level policy to select a behavior population, and add new behavior during deployment.

2.4 Significance and Usefulness

The problem the authors described is important since the limitation of previous work was not to handle continuous environment where some changes needed to be considered can happen during deploying agents in real-world problem. In continuous environment, it is hard to observe the influence of each agent's action and desirable optimal team performance.

2.5 Technical Soundness

The proposed R-MFL approach is sound. It builds on the existing Multi-Fitness Learning (MFL) framework, addressing limitations such as adaptability to environmental changes and catastrophic forgetting. The hierarchic structure of behaviors and policies combined with value iteration guarantees robustness and flexibility. Figure 3 in the paper shows an agent identifying and reacting to a drop in global reward, restoring performance effectively. Given the controlled experiments and comparisons against well-established baselines, the results are convincing. I would like to see evaluation under different reward sparsity levels and more complex environmental dynamics.

2.6 Analysis, Impact, and Conclusion

The authors provide a sound analysis for their results. They provide clear demonstration of cause-and-effect such as the impact of introducing new behaviors on both local and global rewards. From the results, the comparison between R-MFL and baseline methods like MFL is well-executed, isoloating key differences in adaptation strategies. The implications in the paper are not clear. The author explained the results of experiment in detail, not providing the inner-meaning of their experiments. They highlighted the impact of their work. For long-term autonomous operations, robustness and adaptability in multi-agent systems is critical. Moreover, they outline future works in the conclusion part such as combining R-MFL with locally optimal controllers for more efficient performance.

2.7 Background and Reference

The background section in this paper does not explicitly exist. However, in the related works section, the authors sufficiently enumerated learning algorithms and architectures that have been considered for multi-robot deployments and explains why these methods are not effective to handle specific problems the authors mentioned. The references were useful for showing previous works and supporting the authors' points. The number of references is 31, which is sufficient.