Policy Transfer Methods in RoboCup Keep-Away

Extended Abstract

Sabre Didi
Department of Computer Science
Cape Town, South Africa
sabredd0@gmail.com

Geoff Nitschke
Department of Computer Science
Cape Town, South Africa
gnitschke@cs.uct.ac.za

KEYWORDS

Transfer Learning, Reinforcement Learning, HyperNEAT, Novelty Search, RoboCup Soccer

ACM Reference Format:
Sabre Didi and Geoff Nitschke. 2018. Policy Transfer Methods in RoboCup Keep-Away: Extended Abstract. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Kyoto, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3205651.3205710

1 INTRODUCTION

Policy (behavior) transfer is a method to speed-up and improve learning by leveraging knowledge from learning in related but simpler tasks. That is, learned information is reused and shared between a *source* and *target* tasks, where target tasks were used as a starting point for continuing learning [16]. Policy transfer has been widely studied in the context of *Reinforcement Learning* (RL) methods [21], where various studies have consistently demonstrated that transferring knowledge learned on a source task accelerates learning and increases solution quality in target tasks by exploiting relevant prior knowledge [22].

Policy transfer used in company with various RL methods has boosted solution quality in various single-agent tasks including pole-balancing [1], game-playing [17], robot navigation as well as multi-agent tasks including predator-prey [2]. For such single and multi-agent tasks, policy transfer is typically done within the same task domain for varying task complexity [25]. Recently, policy transfer has been used in company with *Evolutionary Algorithms* (EAs) [6] to boost evolved solution quality of evolved genotypes with various representations across various tasks. For example, extracting behavioral features to shape rewards in evolving robot neural controllers for increasingly complex ball collecting tasks [5], and evolving groups of robot neural controllers for navigation tasks that were then used as evolutionary starting points for adaptation in other navigation tasks with different objectives [13]. However, *Neuro-Evolution* (NE) [7] for adapting multi-agent behaviors given policy transfer has received little attention with a few exceptions [26], [4], [15].

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07…$15.00
https://doi.org/10.1145/3205651.3205710

Previous work on *evolutionary policy transfer* has only used fitness functions [6] (*objective-based search*) to direct multi-agent behavior evolution. Thus, the impact of *non-objective* search methods such as behavioral diversity maintenance [14] used in company with policy transfer remains unknown. That is, previous work has only tested single agent tasks such as robot navigation [13] and object collection [5] and simple multi-agent tasks using few agents [24], [26], where non-objective evolutionary search methods were not considered. Furthermore, there has been little research that compares the efficacy of NE versus RL for multi-agent behavior adaptation coupled with policy transfer for boosting solution multi-agent behavior quality across increasingly complex tasks [20], [28].

Non-objective search methods [14] such as *novelty search* [12], out-perform objective based search in various control tasks defined by complex, high dimensional and deceptive fitness landscapes [3], [10]. However, recent results suggest that neither objective or non-objective based search perform optimally when applied to evolve controllers to solve complex multi-agent tasks [4]. Rather, hybridizing objective and non-objective search facilitates the evolution of high quality behaviors [11], [8], [9].

This study investigates multi-agent policy transfer coupled with behavior adaptation by objective and non-objective search variants of HyperNEAT [18] in *RoboCup keep-away* [23]. For comparison, evolved behaviors were compared to those adapted by RL methods: *SARSA* [21] and *Q-Learning* [27], coupled with policy transfer. Keep-away was selected as it is an established multi-agent experimental platform [23]. Similarly, the SARSA and Q-Learning methods were selected as both have been demonstrated for boosting behavior quality with policy transfer [22]. Keep-away behaviors were gauged in terms of *effectiveness* and *efficiency*. Effectiveness was average task performance given policy transfer, where task performance was average ball control time by the keeper team. Efficiency was average number of evaluations taken to reach a minimum task performance threshold given policy transfer.

Research objectives were derived given previous policy transfer research [15][20, 23, 28] .

(1) Demonstrate that keep-away behaviors evolved using hybridized novelty and objective-based search, consistently boosts evolved behavior quality and efficiency, compared to pure novelty or objective-based search.
(2) Demonstrate that such keep-away behaviors evolved using hybridized evolutionary search and policy transfer outperform RL methods in terms of quality and efficiency.

This study's contribution was to elucidate that hybrid evolutionary search out-performs RL methods that have traditionally been task performance benchmarks for policy transfer.

2 METHODS

Keep-away behaviors were adapted using NE or RL. That is, HyperNEAT [18], *SARSA* [21] or *Q-Learning* [27]. Keep-away behavior was first adapted in the source task (*3vs2* keep-away) and then transferred to more complex target tasks (*4vs3*, *5vs4*, *6vs5* keep-away) for further adaptation, where *KvsT* was the number of keeper (*K*) versus taker (*T*) agents, respectively. These methods were selected to elucidate how policy transfer impacts HyperNEAT compared to the RL methods, in terms of average adaptation efficiency and task performance. HyperNEAT-BEV [26] was used to facilitate policy transfer as it has been demonstrated as an effective method for evolved keep-away (multi-agent) policy transfer [15].

Given previous work [19, 20, 23], SARSA [21] and Q-Learning [27] were applied for learning keep-away policies, where the learning goal was for homogeneous keeper teams to select action sequences maximizing total long term reward and thus episode length [28]. To facilitate RL multi-agent policy transfer from the source to a target task a vector of weights, associated with each feature set was periodically stored in memory at 150 episode intervals (equivalent to one NE run). The RL policy transfer function extended *Transfer via Inter-Task Mapping* [23], where weights for source task features from the final episode of source task learning were extracted for transfer to a target task.

3 EXPERIMENTS AND RESULTS

In the HyperNEAT experiments, keep-away policies were evolved for 30 generations in the source task, transferred to a target task and further evolved for another 70 generations. Each generation was 30 episodes and each episode comprised 4500 iterations and tested random initial keeper and taker positions. For comparison, keep-away behavior was also evolved *from scratch* for 100 generations. In RL experiments, keep-away policies were learned over 4500 episodes (equivalent to 30 generations and population size of 150 in HyperNEAT) in the source task. Policies were then transferred to a target task and further adapted for 10500 episodes (70 generations and population size of 150 in HyperNEAT). For comparison, non-policy transfer RL experiments were also run (15000 episodes).

For all tasks, hybrid evolutionary search coupled with policy transfer yielded significantly higher average task performances compared to objective and novelty search variants and RL methods. The overall effectiveness of the hybrid search was consequent of beneficial interactions between behavioral diversity maintenance and objective based evolutionary search [15]. That is, behavioral diversity maintenance first covered large behavior space regions and then in such diverse behavior regions objective-based search was a fine tuning mechanism, following fitness gradients to propagate the fittest keep-away behaviors overall [15].

Thus, results supported the efficacy of hybrid novelty-objective based search for evolving effective behaviors (compared to other HyperNEAT search variants and RL methods) across increasingly complex keep-away tasks when coupled with policy transfer. However, RL methods yielded significantly higher efficiency when coupled with policy transfer but yielded significantly lower effectiveness compared to all tested HyperNEAT search variants.

This study's contribution was to support the benefits of hybrid objective-novelty HyperNEAT search when coupled with policy transfer for boosting the effectiveness of evolved multi-agent behaviors given increasing task complexity. Also, results contributed to increasing empirical evidence supporting the effectiveness of hybridized evolutionary search in complex tasks [14], [9].

REFERENCES

[1] H. Ammar, K. Tuyls, M. Taylor, K. Driessens, and G. Weiss. 2012. Reinforcement Learning Transfer via Sparse Coding. In *Proceedings of the eleventh international conference on autonomous agents and multiagent systems*. AAAI, Spain, 4–8.

[2] G. Boutsioukis, I. Partalas, and I. Vlahavas. 2012. Transfer Learning in Multi-Agent Reinforcement Learning Domains. In *Recent Advances in Reinforcement Learning*. Springer, 249–260.

[3] A. Cully, J. Clune, D. Tarapore, and J. Mouret. 2015. Robots that can adapt like animals. *Nature* 521(1) (2015), 503–507.

[4] S. Didi and G. Nitschke. 2016. Hybridizing Novelty Search for Transfer Learning. In *Proceedings of IEEE Symposium Series on Computational Intelligence*. IEEE, Athens, Greece, 10–18.

[5] S. Doncleux. 2014. Knowledge Extraction from Learning Traces in Continuous Domains. In *AAAI Symposium on Knowledge, Skill, and Behavior Transfer in Autonomous Robots*. AAAI Press, 1–8.

[6] A. Eiben and J. Smith. 2003. *Introduction to Evolutionary Computing*. Springer-Verlag, Berlin, Germany.

[7] D. Floreano, P. Dürr, and C. Mattiussi. 2008. Neuroevolution: from architectures to learning. *Evolutionary Intelligence* 1, 1 (2008), 47–62.

[8] J. Gomes and A. Christensen. 2013. Generic behavior similarity measures for evolutionary swarm robotics. In *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 199–206.

[9] J. Gomes, P. Mariano, and A. Christensen. 2015. Devising Effective Novelty Search Algorithms: A Comprehensive Empirical Study. In *Proceedings of the Genetic Evolutionary Computation Conference*. ACM, Madrid, Spain, 943–950.

[10] J. Gomes, P. Mariano, and A. Christensen. 2016. Novelty-driven Cooperative Coevolution. *Evolutionary Computation* 25(2) (2016), 275–307.

[11] J. Gomes, P. Urbano, and A. Christensen. 2013. Evolution of swarm robotics systems with novelty search. *Swarm Intelligence* 7 (2013), 115–144.

[12] J. Lehman and K. Stanley. 2011. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary computation* 19(2) (2011), 189–223.

[13] A. Moshaiov and A. Tal. 2014. Family Bootstrapping: A Genetic Transfer Learning Approach for Onsetting the evolution for a Set of Realated Robotic Tasks. In *Proceedings of the Congress on Evolutionary Computation*. IEEE Press, 2801–2808.

[14] J. Mouret and S. Doncieux. 2012. Encouraging Behavioral Diversity in Evolutionary Robotics: An Empirical Study. *Evolutionary Computation* 20(1) (2012), 91–133.

[15] G. Nitschke and S. Didi. 2017. Behavior Transfer and Evolutionary Search in Collective Robotics. *Frontiers in Robotics and AI | Evolutionary Robotics* 4(62) (2017), 1–25.

[16] S. Pan and Q. Yang. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10) (2010), 1345–1359.

[17] J. Ramon, K. Driessens, and T. Croonenborghs. 2007. Transfer learning in reinforcement learning problems through partial recycling. In *Proceedings of the 18th European Conference on Machine Learning*. Springer, Warsaw, Poland, 699–707.

[18] K. Stanley, D. D'Ambrosio, and J. Gauci. 2009. A Hypercube-Based Indirect Encoding for evolving large-scale neural networks. *Artificial Life* 15, 2 (2009), 185–212.

[19] P. Stone and R. Sutton. 2002. Keepaway Soccer: A Machine Learning Testbed. In *RoboCup-2001: Robot Soccer World Cup V*. Springer, Berlin, Germany, 214–223.

[20] P. Stone, R. Sutton, and G. Kuhlmann. 2006. Reinforcement learning for RoboCup-soccer keepaway. *Adaptive Behavior* 13(3) (2006), 165–188.

[21] R. Sutton. 1998. Learning to Predict by Methods of Temporal Difference. *Machine Learning* 3, 1 (1998), 9–44.

[22] M. Taylor and P. Stone. 2009. Transfer Learning for Reinforcement Learning Domains: A survey. *Journal of Machine Learning Research* 10(1) (2009), 1633–1685.

[23] M. Taylor, P. Stone, and Y. Liu. 2010. Transfer learning via inter-task mappings for temporal difference learning. *Journal of Machine Learning* 8(1) (2010), 2125–2167.

[24] M. Taylor, S. Whiteson, and P. Stone. 2006. Transfer Learning for Policy Search Methods. In *ICML 2006: Proceedings of the Twenty-Third International Conference on Machine Learning Transfer Learning Workshop*. ACM, Pittsburgh, USA, 1–4.

[25] L. Torrey and J. Shavlik. 2009. Transfer Learning. In *Handbook of Research on Machine Learning Applications*. IGI Global, 17–23.

[26] P. Verbancsics and K. Stanley. 2010. Evolving static representations for task transfer. *Journal of Machine Learning Research* 11, 1 (2010), 1737–1763.

[27] C. Watkins. 1989. *Learning from Delayed Rewards. PhD Thesis*. University of Cambridge, UK.

[28] S. Whiteson and P. Stone. 2006. Evolutionary function approximation for reinforcement learning. *Journal of Machine Learning Research* 7(1) (2006), 877–917.