Neuro-Evolution for Multi-Agent Policy Transfer in RoboCup Keep-Away

(Extended Abstract)

Sabre Didi, Geoff Nitschke
Department of Computer Science
University of Cape Town
Cape Town, South Africa
sabredd0@gmail.com, gnitschke@cs.uct.ac.za

ABSTRACT

An objective of transfer learning is to improve and speed-up learning on target tasks after training on a different, but related source tasks. This research is a study of comparative *Neuro-Evolution* (NE) methods for transferring evolved multi-agent policies (behaviors) between multi-agent tasks of varying complexity. The efficacy of five variants of two NE methods are compared for multi-agent policy transfer. The NE method variants include using the original versions (search directed by a fitness function), behavioural and genotypic diversity based search to replace objective based search (fitness functions) as well as hybrid objective and diversity (behavioral and genotypic) maintenance based search approaches. The goal of testing these variants to direct policy search is to ascertain an appropriate method for boosting the task performance of transferred multi-agent behaviours. Results indicate that an indirect encoding NE method using hybridized objective based search and behavioral diversity maintenance yields significantly improved task performance for policy transfer between multi-agent tasks of increasing complexity. Comparatively, NE methods not using behavioral diversity maintenance to direct policy search performed relatively poorly in terms of efficiency (evolution times) and quality of solutions in target tasks.

Keywords

Machine Learning; Policy Transfer; Neuro-Evolution

Neuro-Evolution for Policy Transfer

This study presents a comparative evaluation of various neuro-evolution methods for multi-agent policy (behavior) transfer, where *Keep-Away RoboCup Soccer* is the experimental case study. This study ascertains the most appropriate method for transfer between tasks of increasing complexity.

We use the *Neuro-Evolution for Augmenting Topologies* (NEAT) [6] and HyperNEAT [5] methods for transfer learning in multi-agent keep-away tasks of varying complexity. These methods were selected in order to test the efficacy of a direct encoding (NEAT) versus an indirect encoding method (HyperNEAT) for multi-agent policy transfer, where such methods have been widely demonstrated as effective for controller design in various multi-agent tasks [8], [1]. Whilst many studies support the efficacy of objective-based search approaches in transfer learning [9], [7], [8], the impact of *genotypic* and *behavioral diversity maintenance* on transfer learning remains unexplored.

Five variants of both NEAT and HyperNEAT for directing the policy search process were tested. *Variant 1* tests unmodified versions of these methods. In *variant 2* behavioral diversity maintenance (*Novelty Search* [3]) replaced the objective (fitness) function. *Variant 3* used a hybrid of objective based search and novelty search. *Variant 4* used a hybrid of objective based search and genotypic diversity maintenance. In *Variant 5* genotypic diversity maintenance replaced the fitness function.

This study investigates how behavioral and genotypic diversity maintenance, non-objective and objective search impacts policy transfer using direct (NEAT) and indirect encoding (HyperNEAT) methods to evolve behaviors. Also, these methods extend previous work on *inter-task mappings for policy search* [7] to facilitate transfer learning [8].

Hypothesis 1 is that, given related policy transfer results [8], NEAT and HyperNEAT are appropriate policy (multi-agent behavior) search methods for enabling policy transfer where transferred behaviors yield a significantly higher task performance and efficiency compared to those without policy transfer (*evolved from scratch*). Hypothesis 2 is that, given behavioral diversity maintenance results [4], [2], if novelty search is hybridized with objective based search in the tested policy search methods, this will yield a significantly higher *task performance* for all source and target tasks, compared to the other method variants tested.

Experiments

Experiments were run in a source keep-away task, where populations evolved after 20 generations (using NEAT or HyperNEAT), were transferred to a target task and evolved for a further 50 generations. Results were compared to those where no policy transfer took place, but rather where NEAT or HyperNEAT evolved keep-away behaviours from *scratch* in target tasks. For both NEAT and HyperNEAT, each genotype (agent team) was evaluated over 30 task trials per generation. Each task trial tested different (random) agent positions. Average fitness per genotype was computed over these 30 task trials. Policy transfer occurred between a

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

Experiment	4vs3 Keep-Away	5vs3 Keep-Away	6vs4 Keep-Away
No Policy Transfer			
NEAT	0.438 (0.037)	0.473 (0.052)	0.419 (0.057)
HyperNEAT	0.587 (0.059)	0.765 (0.050)	0.533 (0.044)
Variant 1			
Fitness Policy Transfer			
NEAT	0.482 (0.059)	0.580 (0.069)	0.464 (0.033)
HyperNEAT	0.729 (0.089)	0.873 (0.089)	0.632 (0.038)
Variant 2			
NS Policy Transfer			
NEAT	0.470 (0.03)	0.505 (0.039)	0.460 (0.03)
HyperNEAT	0.707 (0.027)	0.827 (0.033)	0.605 (0.024)
Variant 3			
Fitness + NS Policy Transfer			
NEAT	0.545 (0.047)	0.638 (0.0048)	0.520 (0.036)
HyperNEAT	0.752 (0.054)	0.943 (0.029)	0.697 (0.032)
Variant 4			
Fitness + GD Policy Transfer			
NEAT	0.442 (0.012)	0.456 (0.018)	0.436 (0.016)
HyperNEAT	0.482 (0.041)	0.509 (0.039)	0.468 (0.035)
Variant 5			
GD Policy Transfer			
NEAT	0.432 (0.018)	0.453 (0.029)	0.426 (0.025)
HyperNEAT	0.475 (0.038)	0.497 (0.029)	0.452 (0.025)

Table 1: Average normalized maximum fitness (over 20 runs) for the three experimental setups. Values are portions of the maximum possible hold time (possession of the ball) for the team of keepers. *NS:* Novelty Search. *GD:* Genotype Diversity. Standard deviations are shown in parentheses.

source and incrementally complex target tasks. The source task was three keepers versus two takers (*3vs2*) in a 20 x 20 virtual field[1]. Evolved behaviors (policies) were transferred (and evolution continued) in one of three target tasks, four keepers versus three takers (*4vs3*), five keepers versus three takers (*5vs3*) or six keepers versus four takers (*6vs4*).

Table 1 presents the average normalized maximum fitness attained by teams evolved with no policy transfer and each policy transfer variant. These results have important implications for current transfer learning research, specifically multi-agent policy transfer where neuro-evolution is used to evolve multi-agent behaviors in source and target tasks. Supported by related research [8], results indicated significant *task performance* benefits (Mann-Whitney U test, p-value < 0.05) of policy transfer for increasingly complex versions of *Keep-away RoboCup Soccer*. Also, results supported the efficacy of using an objective-novelty search hybrid (highest performing variant with statistical significance) to direct NEAT and HyperNEAT behavior evolution for policy transfer. This is similarly supported by previous research demonstrating the benefits of hybrid objective-novelty search approaches over pure novelty search [2].

Future work will compare neuro-evolution methods with well established *reinforcement learning* methods in more complex keep-away soccer tasks, as well as testing these methods for policy transfer between different but related multi-agent tasks such as keep-away to multi-agent predator-prey [1].

[1]All experiments were run in *RoboCup Keep-Away version 6* [7]. Source code and executables can be found at: http://people.cs.uct.ac.za/~gnitschke/KeepAway/

Acknowledgements

This research was funded by a PhD Research Fellowship from the Science Faculty at the University of Cape Town and the National Research Foundation of South Africa (NRF).

REFERENCES

[1] D. D'Ambrosio and K. Stanley. Scalable multiagent learning through indirect encoding of policy geometry. *Evolutionary Intelligence Journal*, 6(1):1–26, 2013.

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

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

[4] J. Mouret and S. Doncieux. Encouraging behavioral diversity in evolutionary robotics: An empirical study. *Evolutionary Computation*, 20(1):91–133, 2012.

[5] K. Stanley, D. D'Ambrosio, and J. Gauci. A hypercube-based indirect encoding for evolving large-scale neural networks. *Artificial Life*, 15(2):185–212, 2009.

[6] K. Stanley and R. Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary Computation*, 10(2):99–127, 2002.

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

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

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