# Towards a General Transfer Approach for Policy-Value Networks

Dennis J.N.J. Soemers[1], Vegard Mella[2], Éric Piette[3], Matthew Stephenson[4], Cameron Browne[1], and Olivier Teytaud[2]

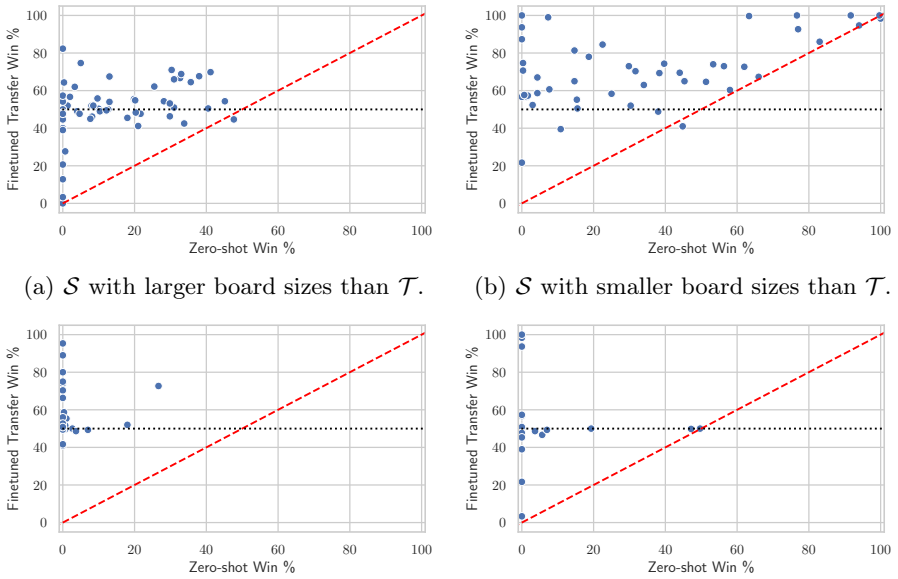
[1] Department of Advanced Computing Sciences, Maastricht University
dennis.soemers@maastrichtuniversity.nl, cambolbro@gmail.com
[2] Meta AI Research vegard.mella@gmail.com, oteytaud@meta.com
[3] ICTEAM, UCLouvain eric.piette@uclouvain.be
[4] College of Science and Engineering, Flinders University
matthew.stephenson@flinders.edu.au

*This document is an encore abstract of the paper entitled "Towards a General Transfer Approach for Policy-Value Networks" [7].*

**Introduction.** AlphaGo Zero [6] and AlphaZero [5] have inspired a successful line of research where policy-value networks—neural networks that have a policy head to output probability distributions over actions for input states, as well as a value head to output state value estimates—for game playing are trained from self-play. Transfer learning [8, 2] may save computation time by transferring such trained networks from some games to others, rather than training them from scratch for every new game. However, prior work on transfer of policy and/or value networks tends to permit very little, if any, variation in the shapes of state or action spaces between the source and target domains. Leveraging the domain specific language (DSL) in which Ludii [3] describes a wide variety of over 1000 distinct board games, we propose a simple baseline transfer approach that can handle transfer between domains that have different state and action spaces.

**Transfer Approach.** We use fully convolutional architectures with global pooling (for the value head) for their ability to handle differences in the *spatial* aspects of state and action spaces (i.e., changes in sizes or shapes of game boards) [4, 9, 1]. Any other differences (e.g., channels encoding presence of piece types that differ between source and target domains) are handled by heuristic rules identifying approximate equivalence relations. These heuristic rules were handcrafted for the entire Ludii system (containing over 1000 board games, which is easily extensible thanks to its DSL) as a whole. They were not handcrafted at the level of individual (pairs of) games.

**Experiments.** We evaluate transfer performance for 150 pairs of variants of games, each of which was used for two transfer experiments (one in either direction). Every pair consists of two variants of the same game (from a pool of nine different board games), with differences in e.g. board sizes, board shapes, or win conditions. We trained a model dedicated to every domain (i.e., game variant)




(a) $\mathcal{S}$ with larger board sizes than $\mathcal{T}$.

(b) $\mathcal{S}$ with smaller board sizes than $\mathcal{T}$.

(c) Different board shapes for $\mathcal{S}$ and $\mathcal{T}$.

(d) Different win conditions for $\mathcal{S}$ and $\mathcal{T}$.

**Fig. 1.** Win percentages of models trained on $\mathcal{S}$ and subsequently fine-tuned on $\mathcal{T}$, against models trained only on $\mathcal{T}$—evaluated on $\mathcal{T}$. Every data point is a different $(\mathcal{S}, \mathcal{T})$ pair. The $x$-axis shows the win percentage of the transferred model *without* fine-tuning (i.e., zero-shot performance).

for 20 hours on 8 GPUs and 80 CPU cores. We evaluated the playing strength of every model when transferring it from the source domain $\mathcal{S}$ it was trained on, by matching it up against the model that was trained directly on the target domain $\mathcal{T}$, both without any fine-tuning (zero-shot transfer), and with an additional 20 hours of fine-tuning time on $\mathcal{T}$. In Fig. 1, data points that are further towards the right indicate more successful zero-shot transfer (where anything that is substantially greater than 0% may be argued to be some degree of success). Results below the black, dotted $y = 50\%$ line indicate negative transfer [10], whereas those above it indicate beneficial transfer. Results below the red $y = x$ line indicate that fine-tuning degraded performance relative to the zero-shot transfer performance, whereas results above that line indicate additional benefits from fine-tuning. While there are cases of negative transfer, we also find many cases of successful zero-shot transfer as well as fine-tuning, across a substantially larger and more varied set of pairings of games than prior work.

**Acknowledgments.** This work was partially supported by the European Research Council as part of the Digital Ludeme Project (ERC Consolidator Grant #771292).

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References


1. Cazenave, T., Chen, Y.C., Chen, G., Chen, S.Y., Chiu, X.D., Dehos, J., Elsa, M., Gong, Q., Hu, H., Khalidov, V., Li, C.L., Lin, H.I., Lin, Y.J., Martinet, X., Mella, V., Rapin, J., Roziere, B., Synnaeve, G., Teytaud, F., Teytaud, O., Ye, S.C., Ye, Y.J., Yen, S.J., Zagoruyko, S.: Polygames: Improved zero learning. ICGA Journal **42**(4), 244–256 (2020)
2. Lazaric, A.: Transfer in reinforcement learning: a framework and a survey. In: Wiering, M., van Otterlo, M. (eds.) Reinforcement Learning. Adaptation, Learning, and Optimization, vol. 12, pp. 143–173. Springer, Berlin, Heidelberg (2012)
3. Piette, É., Soemers, D.J.N.J., Stephenson, M., Sironi, C.F., Winands, M.H.M., Browne, C.: Ludii – the ludemic general game system. In: Giacomo, G.D., Catala, A., Dilkina, B., Milano, M., Barro, S., Bugarín, A., Lang, J. (eds.) Proceedings of the 24th European Conference on Artificial Intelligence (ECAI 2020). Frontiers in Artificial Intelligence and Applications, vol. 325, pp. 411–418. IOS Press (2020)
4. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(4), 640–651 (2017)
5. Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., Hassabis, D.: A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science **362**(6419), 1140–1144 (2018)
6. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., Hassabis, D.: Mastering the game of Go without human knowledge. Nature **550**, 354–359 (2017)
7. Soemers, D.J.N.J., Mella, V., Piette, É., Stephenson, M., Browne, C., Teytaud, O.: Towards a general transfer approach for policy-value networks. Transactions on Machine Learning Research (2023)
8. Taylor, M.E., Stone, P.: Transfer learning for reinforcement learning domains: A survey. In: Mahadevan, S. (ed.) Journal of Machine Learning Research. vol. 10, pp. 1633–1685 (2009)
9. Wu, D.J.: Accelerating self-play learning in Go. https://arxiv.org/abs/1902.10565v3 (2019)
10. Zhang, W., Deng, L., Zhang, L., Wu, D.: Overcoming negative transfer: A survey. IEEE/CAA Journal of Automatica Sinica **10**(2), 305–329 (2023)