Bootstrap Learning in Autonomous Intelligent Machines and Systems through Transfer Learning.

Ciaran Johnson

January 2022

1 Topic statement

Yu et. al [10] state that despite the promising ability of RL agents to automatically acquire behaviour, the data intensive nature of current approaches precludes the ability to learn multiple tasks, if all tasks are to be learnt individually from scratch. Several fields of Machine Learning (ML) exist which aim to bootstrap learning, two which are of particular interest are: Transfer Learning (TL) and Multitask Learning (MTL). TL focuses on utilising knowledge gained in a source domain to improve the performance of a target learner in a different but similar target domain [14]. MTL on the other hand aims to improve generalisation through an inductive transfer mechanism [3]. Both of these approaches require what is known as positive transfer and suffer from an issue known as Negative Transfer (NT) also refer to as Negative Interference in MTL [12]. NT occurs when utilising knowledge gained in the source domain negatively impacts the ability to learn in the target domain. Similarly, Negative Interference occurs when knowledge previously learned solving one task negatively affects the ability to learn a new task.

Traditionally RL models require a large dataset of interactions with a chosen environment in order to effectively learn a policy or value function. However due to the time constraints placed on many AIMS, in order for RL to be effective they must learn new tasks from relatively few samples. Utilising prior knowledge may be one path to reducing the experience required to learn. This knowledge can come from the target AIMS itself or learning can be distributed across many AIMS that use a variety of sensors and actuators to interact with the environment. AIMS with different sensors and actuators will have different domains, as how they observe the environment and their ability to interact with it will vary greatly. In order to collaborate effectively TL must be utilised to communicate knowledge learnt in the source domain to the target domain. We may achieve this by exploiting techniques such as domain similarity estimation and distant transfer to help mitigate NT. Solving the problem of NT would improve the ability of distributed AIMS to work together and learn from each others interactions, reducing the amount of experience a single AIMS needs to collect in order to act effectively in a given environment. This area will likely build upon that of collaborative Multi-agent RL.

TL to bootstrap learning could also be beneficial in the field of Embodied Intelligence. Embodied Intelligence addresses the question of how to combine the mind (AI) and the body (robotics) together to work effectively in uncertain environments. This requires rich sensory data and the ability to process large amounts of multi-modal data on edge devices due to the time constraints placed on AIMS. As resource constraints diminish through technological developments, the iterative process of building AIMS with additional improved sensors and actuators will continue. Both of these improvements will alter the domains in which the AIMS operates. It would therefore be wasteful to start the learning process from scratch due to such improvements. Consequently, transferring knowledge gained from one AIMS to another would be of great benefit to Embodied Intelligence. TL may be applicable in this area, though if the new AIMS is too far removed from that of the previous iterations then TL could result in NT, reducing the ability of the AIMS to learn new tasks.

2 Research Goal

My research goal is to bootstrap the learning process of AIMS that interact in uncertain environments by utilising prior knowledge gained from different AIMS. Zhang et al. [12] state that ensuring positive transfer in challenging open environments with continual streams of heterogeneous data is an open area of research in the field of TL. As such I have decided to address this challenge during my PhD. S. J. Pan and Q. Yang [8] categorises the three key challenges in Transfer Learning as what to transfer, how to transfer and when to transfer. The first two challenges have been the subject of extensive research with the third challenge of when to transfer only recently gaining traction in the research community. One of the issues that must be addressed in when to transfer is that of NT. The issue of NT will likely arise during the course of my PhD and is an area I wish to contribute to in the current research community.

3 Review of the literature

3.1 Negative Transfer

Zhang et al. [12] in their recent survey of NT pull together research from Rosenstein et al [9] and Pan et al. [8] to establish a set of basic assumptions for TL which when violated can result in NT. These are that the tasks learnt in both domains must be similar; both domains must have a similar data distribution; and there exists a suitable model which can be applied to both domains. They also categorise the approaches used to handle NT in four categories: secure transfer, domain similarity estimation, distant transfer and NT mitigation.

Domain estimation concerns itself with computing the transferability between the source and target domain. If domain similarity is low, we can refuse to transfer knowledge or opt to use distant transfer. Distant TL, also called transitive TL, bridges dramatically different source and target domains through one or more intermediate domains. When it is of a medium similarity, NT mitigation can be applied. When similarity is high, data from both domains can be concatenated and used to learn in the target domain.

Zhu et. al [13] specifies that one issue of TL in RL is Modeling Transferability which addresses whether or to what extent can the knowledge gained from solving one task be applied to solving another. This is akin to domain similarity estimation in TL and therefore solutions used to perform domain similarity estimation in DL models may be applicable to Model Transferability in RL. The domain similarity in DL can be estimated from fine tuning which is frequently used in deep TL to adapt the source domain DL model to the target domain. The model can be fine-tuned on downstream tasks by fixing lower layer parameters and fine-tuning the higher layer parameter.

Multi-task learning solves multiple learning tasks simultaneously by exploiting their commonalities and differences. Similar to TL, it needs positive transfer among tasks to improve the overall learning performance. Previous studies [7, 4] have observed that conflicting gradients among different tasks may induce Negative Interference. Yu et al. [11] proposes PCGrad a gradient surgery method for MTL projecting task gradients onto the normal plane of all conflicting tasks. PCGrad was able to improve performance and and efficiency on various multi-task RL and supervised learning challenges. This approach is useful for avoiding Negative Interference in MTL, though due to the lack of conflicting gradients in single-task TL it will likely be of little benefit when avoiding NT when bootstrapping learning.

3.2 Learning from Demonstration

Learning from Demonstration (LfD) looks to use human knowledge to complete tasks, this is achieved by imitating experts. Adama et al. [2] discuss how Transfer Learning in assistive robotics can be effective through the use of expert demonstrators where the source domain is created through watching an expert (e.g. a nurse in a care home) and generating features, primarily through vision. The knowledge gained through observing the expert is then transferred to the target domain in which the assistive robot operates. Generative Adversarial Imitation Learning (GAIL) [5] is a model-free imitation learning approach which scales relatively well in high-dimensional environments. This approach draws inspiration from Imitation Reinforcement Learning and Generative Adversarial Network and requires few expert interactions to learn expert level policies. The drawback of GAIL is that it still has a high sample-complexity during training meaning it requires an excessive number of interactions with the environment to learn. InfoGAIL [6] is an eXplainable RL (XRL) model which builds upon GAIL by identifying the salient latent factors that differ between demonstrations. These salient latent factors can be useful in interpreting the performance of the model. XRL and eXplainable AI (XAI) is an important field of research gaining increasing attention in research. XAI is not only important for explaining the behaviour of AI models but may also provide an insight in how to transfer knowledge from one domain to another.

However relying on demonstrations from human experts to acquire new task removes a degree of the autonomy from the AIMS. Abbatematteo et al. [1] aim to bootstrap learning using motion planning and therefore avoiding the need for a human demonstrator. Their research uses goal-orientated kinematic planning to generate initial trajectories to reduce sample complexity that can be optimised through policy search. This approach provides a starting point for learning, improving upon learning from scratch with random initialisation yet not outperforming a LfD initialisation where an expert controlled the robot arm's movement through the use of a joystick. Though impressive the paper does not address the issue of transferring knowledge learnt in a source domain to a target domain. Therefore when attempting to bootstrap learning for more complex tasks using knowledge gained during the deployment of one AIMS into a similar yet different AIMS this approach would not be applicable.

4 Research methods and Study Design

The plan for this PhD is to further investigate TL within the fields of RL, MTL, and Embodied Intelligence. Within the field of RL I also intended to focus on Decentralised Cooperative Multi-agent RL, the goal being to discover techniques used in MARL to communicate effective decentralised policies from one agent to another while ensuring positive transfer. I will also explore the different approaches that MTL uses to avoid Negative Interference in order to determine their applicability in TL.

Upon the completion of my PhD I intend to bootstrap learning in a target AIMS by utilising knowledge gained by a source AIMS. This goal can be broken down into a research phase, where previous studies in the fields stated above are collected and collated in an academic paper. Following this research I intend to teach a robot a single task in a fixed deterministic environment, then bootstrap an identical robot so that it can also quickly learn this task. Following the completion of this I will perform multi-task learning on the first device and again use this to bootstrap learning on the second robot. At this stage I will change the robot so that it is similar yet different to the robot which originally learnt both tasks, in essence changing the target domain so that it is no longer equivalent to the source domain. At this stage I must be wary of negative transfer, and given time I will select a robot whose domain drastically differs to the original robot so that NT is induced. I will then attempt to bootstrap this new robot using TL, the knowledge gained from the original robot will be passed to the new robot in order to complete both the single and multitask challenges. Upon successful completion I will deploy the developed methods into an uncertain environment repeating the steps above. This will require the development of new methodologies.

References

[1] Ben Abbatematteo, Eric Rosen, Stefanie Tellex, and George Konidaris. Bootstrapping motor skill learning with motion planning, 2021.

[2] David Adama, Ahmad Lotfi, R. Ranson, and Pedro Trindade. Transfer learning in assistive robotics: From human to robot domain. 01 2019.

[3] Rich Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, July 1997.

[4] Zhao Chen, Jiquan Ngiam, Yanping Huang, Thang Luong, Henrik Kretzschmar, Yuning Chai, and Dragomir Anguelov. Just pick a sign: Optimizing deep multitask models with gradient sign dropout, 2020.

[5] Jonathan Ho and Stefano Ermon. Generative adversarial imitation learning, 2016.

[6] Yunzhu Li, Jiaming Song, and Stefano Ermon. Infogail: Interpretable imitation learning from visual demonstrations, 2017.

[7] Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning, 2021.

[8] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[9] Michael Rosenstein, Zvika Marx, Leslie Kaelbling, and Thomas Dietterich. To transfer or not to transfer. 01 2005.

[10] Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Multi-task reinforcement learning without interference. 2019.

[11] Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning, 2020.

[12] Wen Zhang, Lingfei Deng, Lei Zhang, and Dongrui Wu. A survey on negative transfer, 2021.

[13] Zhuangdi Zhu, Kaixiang Lin, and Jiayu Zhou. Transfer learning in deep reinforcement learning: A survey, 2021.

[14] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *CoRR*, abs/1911.02685, 2019.