

Transfer Learning Progress and Potential

Ted E. Senator

■ *This article relates transfer learning to other AI research areas, envisions several new types of transfer learning capabilities, and suggests several new research directions to achieve them.*

As evidenced by the articles in this special issue, transfer learning has come a long way in the past five or so years, partially because of DARPA's Transfer Learning program, which sponsored much of the work reported in this issue.¹ Work conducted as part of this program and work conducted under other auspices, as evidenced by the many conference and workshop papers,² demonstrates the interest and promise in this approach to building intelligent agents. There is a Transfer Learning Toolkit for Matlab available on the web.³ Transfer learning has developed techniques for classification, regression, and clustering (as summarized in Pan and Yang's 2009 survey) and for complex interactive tasks that are often best addressed by reinforcement learning techniques. And transfer learning has been applied to domains as diverse as named entity recognition, image clustering, information retrieval, link prediction, AP physics, and others.

As with many human-level AI goals, transfer learning is still a long way from the ability for agents to take advantage of relevant previous learned knowledge and experience to perform (at least) competently and effectively on new tasks the first time they are encountered. However, there is a more practical and more feasible goal for transfer learning against which progress is being made. An engineering-oriented goal of artificial intelligence that could be enabled by transfer learning is the ability to construct a large number of diverse applications not from scratch, but by taking advantage of knowledge already acquired and formally represented for other purposes. This would be useful when knowledge acquisition is expensive, when sufficient amounts of training data for a new task are unavailable, or when the new task must be learned more rapidly than possible from directly relevant training data.

As early as 1983, Jaime Carbonell, Ryszard Michalski, and Tom Mitchell (Carbonell, Michalski, and Mitchell 1983) recognized that machine learning was a promising technique for overcoming the knowledge-acquisition bottleneck in the construction of AI applications. In 1991 Robert Neches and colleagues (Neches et al. 1991) suggested knowledge sharing and reuse as another cost-effective method to build large numbers of AI systems. Their article does not, however, mention the word *learning* even once. Combining the ideas of using machine learning for knowledge acquisition and of reusing and sharing previously acquired knowledge leads to the idea of transfer learning; that is, using knowledge learned for one task or domain to more rapidly achieve competence on a new task or domain.

Two other active areas of research could also enable the rapid construction of AI applications that use previously acquired and represented knowledge. The semantic web is attempting to make the large amount of knowledge that has been represented in a form suitable for human consumption available for machine reasoning. Another approach, being pursued by DARPA's Machine Reading program, is to enable the transformation of knowledge represented in naturally occurring text into the formal representations used by AI reasoning systems. Both of these approaches assume that the knowledge needed for a task already exists in human consumable form, either on the web or in text, respectively, in contrast to transfer learning, which assumes the existence of agents that can perform related tasks that may not have been encountered previously.

Let's imagine a general-purpose intelligent agent that is experienced at some tasks and competent at others, but has never seen or attempted the large variety of possible tasks in the world. One can view this agent as a child, learning to function in a complex world, or as an adult, with a broad range of competencies and a small number of specialties. One can even imagine that the agent encounters a novel task or situation that has never before been encountered by it or any other agent, because the world changes and new tasks and domains arise continually. What would this agent have to do to perform competently on this new task? A series of steps might be:

1. Recognize the task — what is this task? Is it something that has been seen before or is it new? What are its characteristics? Note that these questions assume that the agent has available to it some ontology in which the task may be situated, even if only generally.
2. Decompose the task into subtasks that may have been encountered before, even as unrecognized subtasks of previous monolithic tasks.
3. Recognize the task as an instance of a more general task that has been solved before, or as an

instance of a generalization of a similar task that has been solved before.

4. Find similar tasks it has solved before, and understand how they were solved and how well.
5. Identify what the current task shares with the previous tasks, and how it differs.
6. Identify what can be reused from old tasks, what needs to be modified, and what needs to be combined from the related previous tasks.
7. Attempt to solve the new task, using relevant knowledge from previously solved tasks.
8. Figure out how to tell if the new task is solved. (Is there feedback from the environment? Is it in the task specification itself?)
9. Figure out where the gaps in the solution came from.
10. Iterate back to (all) previous steps as needed.

What is missing from today's transfer learning techniques from this list? It seems that the main deficiency is the ability to recognize and understand the original and new tasks in a general context; rather, current transfer learning techniques operate in specific predefined task contexts that have been set up in advance by a researcher. This suggests a potential avenue for new research — a generalized form of transfer learning in which a broad range of agents with experience in many tasks and domains is available and in which the transfer learning agent must figure out from which of them to select and adapt knowledge for the new task. In contrast to current transfer learning techniques that have to figure out only how to modify previous knowledge to perform a new task, such general transfer learning agents would also have to figure out what previous knowledge to select, combine, and modify to perform their new task. Such an agent would be closer to the human ability to apply lessons from previous experience in new situations. One can even imagine that a population of different transfer learning agents could be created in this manner for multiple specialized tasks, providing in the aggregate the ability to act intelligently in a wide range of new situations, given the ability to recognize tasks for which they are competent. Providing agents with the representational mechanisms to make this type of general transfer possible would likely be a key aspect of such research, and providing rich testbeds and environments would be a key enabler.

Another potential direction for future research is the exploration of more formal techniques for specifying the notion of "task relatedness," which is at the core of transfer learning. The recent dissertation by Hassan Mahmud (Mahmud 2008) presents a general measure of relatedness between tasks based on ideas from algorithmic information theory and is a step in this direction. The DARPA transfer learning program began with the idea that the difficulty of transfer learning would increase as the tasks became less related; however, other dimensions such as what to transfer and the trans-



Artificial Intelligence and Interactive Digital Entertainment

**Seventh Annual
Artificial
Intelligence
and Interactive
Digital
Entertainment
Conference
(AIIDE-11)**

Save the Date!
October 11–14, 2011



K- CAP 2011

The Sixth International Conference on Knowledge Capture

Dates: 25 - 29 June 2011

Banff Conference Centre

Banff, Alberta, Canada

Full paper Submission deadline: Feb 15, 2011



<http://kcap11.stanford.edu>

fer technique have been emphasized more as the research progressed.

Combining ideas from transfer learning with other learning techniques also would enable powerful new capabilities. The survey by Taylor and Stone (2009) reviews work that uses transfer learning in complex domains amenable to reinforcement learning. The Office of Naval Research has recently released a Broad Agency Announcement for research in Active Transfer Learning that addresses the causes and potential repairs of failed attempts at transfer learning.

These developments all point to a productive future for further advances in transfer learning as a component of artificially intelligent agents.

Notes

1. See www.darpa.mil/i2o/programs/tl/tl.asp for an overview of the vision for the DARPA Transfer Learning program.
2. See www1.i2r.a-star.edu.sg/~jspan/conferenceTL.htm.
3. See multitask.cs.berkeley.edu.

References

Carbonell, J. G.; Michalski, R. S.; and Mitchell, T. M. 1983. Machine Learning: A Historical and Methodological Analysis. *AI Magazine* 4(3): 69–79.

Defense Advanced Research Projects Agency (DARPA). 2009. Machine Reading Broad Agency Announcement 09-03. Washington, DC: United States Department of

Defense (www.darpa.mil/ipto/solicit/baa/BAA-09-03_PIP.pdf).

Defense Advanced Research Projects Agency (DARPA). 2005. Transfer Learning Proposer Information Pamphlet (PIP) for Broad Agency Announcement 05-29. Washington, DC: United States Department of Defense (www.darpa.mil/ipto/solicit/baa/BAA-05-29_PIP.pdf).

Mahmud, M. M. Hassan. 2008. Universal Transfer Learning. Ph.D. Diss., Department of Computer Science, University of Illinois at Urbana-Champaign, IL.

Neches, R.; Fikes, R.; Finin, T.; Gruber, T.; Patil, R.; Senator, T.; and Swartout, W. R. 1991. Enabling Technology for Knowledge Sharing. *AI Magazine* 12(3): 16–36.

Office of Naval Research. 2010. Broad Agency Announcement (BAA) for Active Transfer Learning 10-017. Washington, DC: United States Navy (www.onr.navy.mil/~media/Files/Funding-Announcements/BAA/10-017.ashx).

Pan, S. J.; and Yang, Q. 2009. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10): 1345–1359 (doi.ieeecomputersociety.org/10.1109/TKDE.2009.191).

Taylor, M. E., and Stone, P. 2009. Transfer Learning for Reinforcement Learning Domains: A Survey. *Journal of Machine Learning Research* 10(2009): 1633–1685.

Ted Senator is a vice president and technical fellow at Science Applications International Corporation (SAIC) (www.saic.com). As a program manager at DARPA, he initiated DARPA's Transfer Learning program. He serves as the secretary-treasurer of AAAI.