Guest editor's introduction: special issue on inductive transfer learning

Daniel L. Silver · Kristin P. Bennett

Received: 1 October 2008 / Accepted: 1 October 2008 / Published online: 21 October 2008
Springer Science+Business Media, LLC 2008

1 Introduction

Inductive transfer or transfer learning refers to the problem of retaining and applying the knowledge learned in one or more tasks to develop efficiently an effective hypothesis for a new task. While all learning involves generalization across problem instances, transfer learning emphasizes the transfer of knowledge across domains, tasks, and distributions that are similar but not the same. For example, learning to recognize chairs might help to recognize tables; or learning to play checkers might improve the learning of chess. While people are adept at inductive transfer, even across widely disparate domains, we have only begun to develop associated computational learning theory and there are few machine learning systems that exhibit knowledge transfer.

Inductive transfer invokes some of the most important questions in artificial intelligence. Amongst its challenges are questions such as:

- What is the best representation and method for retaining learned background knowledge? How does one index into such knowledge?
- What is the best representation and method for transferring prior knowledge to a new task?
- How does the use of prior knowledge affect hypothesis search heuristics?
- What is the nature of similarity or *relatedness* between tasks for the purposes of learning? Can it be measured?
- What role does curriculum play in the sequential learning of tasks?

The papers in this special issue consider several of these problems.

D.L. Silver (✉)

Jodrey School of Computer Science, Acadia University, Wolfville, NS, Canada B4P 2R6
e-mail: danny.silver@acadiau.ca

K.P. Bennett

Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180-3590, USA
e-mail: bennek@rpi.edu

2 Brief history of inductive transfer

Research in inductive transfer began in the early 1980s with discussions on inductive bias, generalization and the necessity of heuristics for developing accurate hypotheses from small numbers of training examples. Mitchell's work in 1980 defined five categories of inductive bias, and amongst these were "analogy with previously learned tasks" and "knowledge of the task domain" (Mitchell 1980). Utgoff and others followed up this work by pointing out that, ideally, a learning system should be able to change its inductive bias to tailor its preference for hypotheses according to the task being learned (Utgoff 1986). The ability to change inductive bias requires that the learning system have prior knowledge about some aspect of the task domain. Furthermore, it suggests that the accumulation of prior knowledge at the inductive bias level is a useful characteristic for any learning system.

From 1986 to 1995 there were a number of important early papers that examined the efficiency of model development and the performance of models as a function of inductive transfer. In 1994, the ML-COLT workshop on "Constructive Induction and Change of Representation" played an important role by examining the many ways of generating or modifying model representations automatically (ML-COLT'94 workshop 1994). Workshop attendees emphasized the futility of bias-free learning and the need to investigate systems that used prior domain knowledge as a source of inductive bias. A good bibliography of papers from this period can be found at NIPS'05 workshop (2005).

The first major meeting on inductive transfer occurred in 1995 at the NIPS workshop on "Learning to Learn" in Vail, Colorado (NIPS'95 workshop 1995). The workshop focused on the need for lifelong machine learning methods that retain and reuse learned knowledge. The co-organizers of that workshop were Rich Caruana, Danny Silver, Jon Baxter, Tom Mitchell, Lorien Pratt, and Sebastian Thrun. The fundamental motivation for this meeting was the acceptance that machine learning systems would benefit from manipulating knowledge learned from related prior experience and that this would enable them to move beyond task-specific *tabula rasa* induction. The 1995 workshop identified the following as the most important areas for future research:

- The relationship between computational learning theory and selective inductive bias;
- The tradeoffs between storing or transferring knowledge in representational and functional form;
- Methods of turning concurrent parallel learning into sequential lifelong learning methods;
- Measuring relatedness between learning tasks for the purpose of knowledge transfer;
- Long-term memory methods and cumulative learning; and
- The practical applications of inductive transfer and lifelong learning systems.

The workshop resulted in a series of articles published in special issues of Connection Science (Pratt 1996) and Machine Learning (Pratt and Thrun 1997), and a book entitled "Learning to Learn" (Thrun and Pratt 1997).

Research in inductive transfer continued after 1995 under a variety of names: bias learning, speedup learning, learning to learn, machine life-long learning, knowledge transfer, transfer learning, multiple task learning, knowledge consolidation, context-sensitive learning, knowledge-based inductive bias, meta-learning, and incremental, cumulative, and continual learning. The years 2003 through 2005 saw a renewed interest in multi-task learning using kernel and Bayesian approaches that established new frameworks for inductive transfer and defining task relatedness. For an extensive list of publications please see NIPS'05 workshop (2005).

Between 1995 and 2005 there were approximately ten workshops or symposia related to inductive transfer. Most notably there was a series of workshops on *meta-learning*, that

began with the 1995 ECML'95 Workshop on Learning at the Knowledge Level. In addition there have been workshops on context-sensitive learning, life-long learning, and learning in *a priori* unknown or dynamic domains. For a complete history of workshops and symposia please see NIPS'05 workshop (2005).

In 2005, the second major workshop on inductive transfer was held. This one-day workshop entitled “Inductive Transfer: 10 Years Later” examined the progress that had been made over the previous ten years, the questions and challenges that remained, and the opportunities for new applications of inductive transfer systems. The workshop was co-organized by Danny Silver, Rich Caruana, Kristin Bennett, Goekhan Bakir, Massimiliano Pontil, Stuart Russell, and Prasad Tadepalli.

In particular, the workshop organizers identified three major goals: (1) To summarize the work in the area of inductive transfer so as to develop a taxonomy of research indicating open questions, (2) To share new theories, approaches and algorithms regarding the accumulation and use of learned knowledge for the purposes of more effective and efficient learning, (3) To discuss a more formal inductive transfer community (or special interest group) that might begin by offering a repository website, benchmarking data and testing methods, shared software, and links to various research programs and other resources. The workshop was also intended to discuss more forward looking questions such as:

- Under what conditions is inductive transfer difficult? When is it easy?
- What are the fundamental requirements for continual learning and transfer?
- What new mathematical models and frameworks capture or demonstrate transfer learning?
- What are the latest and most advanced demonstrations of transfer learning including Bayesian approaches, kernel methods, and reinforcement learning?
- What can be learned from transfer learning in humans and animals?
- What are the latest psychological, neurological, and computational theories of knowledge transfer in learning?

Based on 25 original submissions, the workshop featured eight paper presentations, four invited speakers, two panel discussions and a poster session. For details see NIPS'05 workshop (2005).

3 Overview of this special issue

This issue on inductive transfer developed from the 2005 NIPS workshop. All 2005 workshop authors and invited speakers were invited to submit papers. From the 13 submitted papers five were selected for this special issue. These five papers cover a wide range of recent work in inductive transfer: theory (Ben-David and Schuller Borbely 2008), kernel-SVM (Argyriou et al. 2008), hierarchical Bayesian (Zhang et al. 2008), reinforcement learning (Mehta et al. 2008), and neural networks (Silver et al. 2008).

The paper “Flexible latent variable models for multi-task learning” (Zhang et al. 2008) by Jian Zhang, Zoubin Ghahramani and Yiming Yang presents a Hierarchical Bayesian probabilistic framework for multi-task learning, where related task parameters share a common structure through a set of latent variables. By making statistical assumptions, the framework can be used to support a set of important latent variable models for different multi-task scenarios. The algorithm alternates between unsupervised and supervised steps. Unsupervised learning at the task domain level determines common representation parameters. This is followed by supervised learning at the task level to determine task-specific values for these

parameters. By learning the related tasks jointly, the approach is able to get a better estimation of the shared components and thus achieve a better generalization capability compared to conventional approaches where the learning of each task is carried out independently. The distributional assumption of latent variables can be seen as estimating the density of task parameters and, in this way, provides a statistical explanation of task relatedness. The paper presents efficient algorithms for the empirical Bayes method as well as point estimation. Empirical studies on synthetic and real-world classification datasets show the effectiveness of the proposed models in two evaluation settings: a standard multi-task learning setting and a transfer learning setting.

“Convex multi-task feature learning” (Argyriou et al. 2008) by Andreas Argyriou, Theodoros Evgeniou and Massimiliano Pontil introduces a method for learning sparse representations shared across multiple tasks that can be considered a generalization of single-task 1-norm method of regularization. The method is based on a novel non-convex regularizer which controls the number of learned features common across the tasks such that it both couples the tasks and enforces sparsity of representation. The method is proven to be equivalent to solving a convex optimization problem for which there is an iterative algorithm which converges to an optimal solution, and is therefore the first convex optimization formulation for multi-task feature learning. Similar to the first paper, this algorithm alternately performs a supervised and an unsupervised step, where in the former step it learns task-specific functions and in the latter step it learns common-across-tasks sparse representations for these functions. The authors develop a novel and computationally efficient extension of the algorithm that learns sparse non-linear representations using kernels. Experiments on simulated and real data sets demonstrate that the proposed method leads to shared features common across related tasks and improved model performance relative to learning each task independently. The algorithm can also be used, as a special case, to select those variables that are common across the tasks without learning a specific task.

Shai Ben-David’s and Reba Schuller Borbely’s paper “A notion of task relatedness yielding provable multiple-task learning guarantees” (Ben-David and Schuller Borbely 2008) makes a significant theoretical step forward by formalizing one perspective on the nature of relatedness between tasks for multiple task learning. Prior theory on inductive transfer has treated the notion of relatedness using a functional approach (Baxter 2000) that relies on the assumption that the tasks share a common optimal inductive bias that reduces the learning method’s search space to a common optimal subset of hypotheses. In contrast, Ben-David and Borbely’s approach focuses on the sample generating distributions underlying the learning tasks, and defines task relatedness as an explicit relationship between these distributions. They provide a formal framework for this notion of task relatedness. Consider a set F of transformations $f : X \rightarrow X$ on the input space X . Tasks are considered F -related if, for some fixed probability distribution over $X \times \{0, 1\}$, the data in each of these tasks is generated by applying some $f \in F$ to that distribution. This notion of task relatedness is relevant to a variety of real-life multi-task learning scenarios; for example, the situation where many different sensors collect data for the same classification problem (intrusion detection). In this case, there will be a set of mathematical transformations F such that the data distributions collected by of all sensors are F -related. It also allows the formal derivation of generalization bounds (based on a generalized VC-dimension parameter) that are strictly stronger than the previously known bounds for multi-task learning scenarios. The authors give precise conditions under which the bounds guarantee generalization on the basis of smaller sample sizes.

“Transfer in Variable-Reward Hierarchical Reinforcement Learning” (Mehta et al. 2008) by Neville Mehta, Sriraam Natarajan, Prasad Tadepalli, and Alan Fern considers inductive

transfer in the context of related Reinforcement Learning (RL) tasks. These RL tasks, derived from Semi-Markov Decision Processes (SMDPs), share the same transition dynamics but have different reward functions that are linear in a set of reward features. The focus here is on learning efficiency. The authors formally define the transfer learning problem as developing an efficient algorithm to solve any SMDP drawn from a fixed distribution after experiencing prior related SMDP tasks. The paper introduces an online algorithm to solve this problem called Variable-Reward Reinforcement Learning (VRRL). VRRL compactly stores prior optimal value functions for SMDPs, and uses them to initialize optimally the value function for a new SMDP. The method is generalized to a hierarchical RL setting where the different SMDPs share the same task hierarchy. Empirical studies with VRRL on a domain of real-time strategy games demonstrates that learning efficiency improves through inductive transfer, especially in hierarchical settings where the overall value functions are decomposed into subtask value functions that can be used in different SMDPs.

In the final paper, Danny Silver, Ryan Poirier and Duane Currie present *context-sensitive* multi-task learning, or *csMTL*, in a paper called “Inductive Transfer with Context-Sensitive Neural Networks” (Mehta et al. 2008). *csMTL* is presented as a method of inductive transfer which uses a single output back-propagation neural network and additional contextual inputs for learning multiple tasks. The approach converts the problem of developing hypotheses for multiple tasks to one of developing a single hypothesis for all tasks that are indexed (or biased) by the context inputs. With *csMTL*, all free parameters (connection weights) in the network are shared as compared to only those below a common feature layer, as with a standard MTL network. Testing on seven synthetic and real-world domains of tasks shows that the method can produce hypotheses that are as good as or better than standard MTL neural networks. The authors argue that the reason for this performance improvement is a reduction in the number of effective free parameters in the *csMTL* network brought about by weight update constraints due to the shared output node and the context inputs. Experiments using decision trees and support vector machines provides initial evidence that not all machine learning methods have the ability to use *csMTL* encoded data to affect beneficial inductive transfer.

We trust that the mix of articles collected in this special issue will spark further interest and curiosity in inductive transfer. There is much work to be done in this area in terms of new computational learning theory and the application of existing algorithms and techniques.

We are very grateful to all authors for their submissions to this special issue and for their work in improving and polishing their articles. We would also like to thank the referees for their reviews and helpful comments to authors that have led to improvements in content and format. Finally, we are particularly grateful to the Editor-in-Chief, Foster Provost, for the opportunity to compile this special issue and to the editorial staff of Machine Learning for their kind assistance.

References

- Argyriou, A., Evgeniou, T., & Pontil, M. (2008). Convex multi-task feature learning. *Machine Learning*. doi:10.1007/s10994-007-5040-8.
- Baxter, J. (2000). A model for inductive bias learning. *Journal of Artificial Intelligence Research*, 12, 149–198.
- Ben-David, S., & Schuller Borbely, R. (2008). A notion of task relatedness yielding provable multiple-task learning guarantees. *Machine Learning*. doi:10.1007/s10994-007-5043-5.
- Mehta, N., Natarajan, S., Tadepalli, P., & Fern, A. (2008). Transfer in variable-reward hierarchical reinforcement learning. *Machine Learning*. doi:10.1007/s10994-008-5061-y.

- Mitchell, T. M. (1980). The need for biases in learning generalizations. In: J. W. Shavlik & Th. G. Dietterich (Eds.), *Readings in machine learning* (pp. 184–191).
- ML-COLT'94 workshop (1994). Constructive induction and change of representation. http://home.comcast.net/~tom.fawcett/public_html/CICR/index.html.
- NIPS'95 workshop (1995). Learning to learn. http://iitrl.acadiau.ca/NIPS95_LTL.
- NIPS'05 workshop (2005). Inductive transfer: 10 years later. <http://iitrl.acadiau.ca/itws05/>.
- Pratt, L. (Ed.) (1996). Reuse of neural networks through transfer. *Connection Science*, 8(2).
- Pratt, L., & Thrun, S. (Eds.) (1997). Transfer in inductive systems. *Machine Learning*, 28(1).
- Silver, D., Poirier, R., & Currie, D. (2008). Inductive transfer with context-sensitive neural networks. *Machine Learning*. doi:[10.1007/s10994-008-5088-0](https://doi.org/10.1007/s10994-008-5088-0).
- Thrun, S., & Pratt, L. Y. (Eds.). (1997). *Learning to learn*. Boston: Kluwer Academic.
- Utgoff, P. E. (1986). *Machine learning of inductive bias*. Boston: Kluwer Academic.
- Zhang, J., Ghahramani, Z., & Yang, Y. (2008). Flexible latent variable models for multi-task learning. *Machine Learning*. doi:[10.1007/s10994-008-5050-1](https://doi.org/10.1007/s10994-008-5050-1).