Chapter 11
Transfer Learning

Lisa Torrey
University of Wisconsin, USA

Jude Shavlik
University of Wisconsin, USA

ABSTRACT

Transfer learning is the improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned. While most machine learning algorithms are designed to address single tasks, the development of algorithms that facilitate transfer learning is a topic of ongoing interest in the machine-learning community. This chapter provides an introduction to the goals, settings, and challenges of transfer learning. It surveys current research in this area, giving an overview of the state of the art and outlining the open problems. The survey covers transfer in both inductive learning and reinforcement learning, and discusses the issues of negative transfer and task mapping.

INTRODUCTION

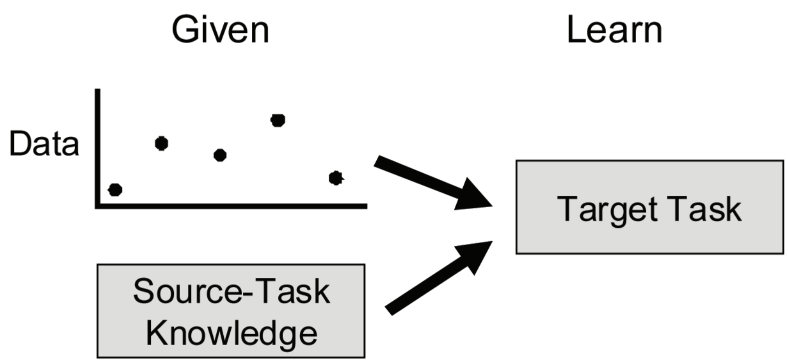
Human learners appear to have inherent ways to transfer knowledge between tasks. That is, we recognize and apply relevant knowledge from previous learning experiences when we encounter new tasks. The more related a new task is to our previous experience, the more easily we can master it.

Common machine learning algorithms, in contrast, traditionally address isolated tasks. Transfer learning attempts to improve on traditional machine learning by transferring knowledge learned in one or more source tasks and using it to improve learning in a related target task (see Figure 1). Techniques that enable knowledge transfer represent progress towards making machine learning as efficient as human learning.

This chapter provides an introduction to the goals, settings, and challenges of transfer learning. It surveys current research in this area, giving an overview of the state of the art and outlining the open problems.

DOI: 10.4018/978-1-60566-766-9.ch011

Figure 1. Transfer learning is machine learning with an additional source of information apart from the standard training data: knowledge from one or more related tasks.



Transfer methods tend to be highly dependent on the machine learning algorithms being used to learn the tasks, and can often simply be considered extensions of those algorithms. Some work in transfer learning is in the context of inductive learning, and involves extending well-known classification and inference algorithms such as neural networks, Bayesian networks, and Markov Logic Networks. Another major area is in the context of reinforcement learning, and involves extending algorithms such as Q-learning and policy search. This chapter surveys these areas separately.

The goal of transfer learning is to improve learning in the target task by leveraging knowledge from the source task. There are three common measures by which transfer might improve learning. First is the initial performance achievable in the target task using only the transferred knowledge, before any further learning is done, compared to the initial performance of an ignorant agent. Second is the amount of time it takes to fully learn the target task given the transferred knowledge compared to the amount of time to learn it from scratch. Third is the final performance level achievable in the target task compared to the final level without transfer. Figure 2 illustrates these three measures.

Figure 2. Three ways in which transfer might improve learning: a higher performance at the very beginning of learning, a steeper slope in the learning curve, or a higher asymptotic performance.