Subsymbolic Re-representation to Facilitate Learning Transfer

Learning transfer is the ability of a system to learn one problem and then to transfer a significant amount of the learned knowledge to a different problem. While the learning transfer itself is often considered a creative act, creativity is additionally required in deciding *which* prior knowledge to use and *how* to use it. Symbolic systems employing some form of analogy have been somewhat successful here; however, these approaches require a significant amount of specialized domain knowledge and do not generalize. We propose the use of sub-symbolic approaches to learning transfer, trading the interpretability of symbolic approaches for representational power and generality.

While previous work has laid the groundwork for sub-symbolic systems that exhibit learning transfer, we propose to build a system capable of *substantive* learning transfer, incorporating a method for measuring the "transferability" of a pair of tasks and a general (sub-symbolic) mechanism for knowledge representation and transfer. This work will focus on the knowledge representation aspect of the system.

To accomplish creative (novel and useful) learning transfer between tasks, acquired knowledge often must be re-represented in some general form. We will employ an "image"-based approach to discover a (re-)representation mechanism that is invariant to various transforms. We consider the general case where closed form analytical expressions for such transforms will not be derivable, and propose learning interesting transforms inherent in the data by employing a neural approach as (a hopefully compact) representation. As a simple example, a system listening to music might discover the concept of transposition and represent it as a neural filter. When new music is experienced, the system can use this neural transposition filter either discriminatively (so the music is recognized invariant to transposition) or generatively (to produce transposed versions of new music).

In many cases, these transformations may occur on a lower-dimensional manifold (that is, lower than the intrinsic representational dimension), and we will have to first discover that surface in order to produce an accurate (re-)representation (in the form of a neural filter). Combining nonlinear manifold learning with a sub-symbolic transform representation will allow us to discover interesting transforms that can be used to (re-)represent data in a way that facilitates learning transfer.

Figure 1 demonstrates the idea. In the process of learning to recognize the letter A, we collect data in the form of examples. The explicit representation of this data (as pixels) may not be an informative representation or it may contain problem specific information that we would like to generalize away. Learning the implicit manifold on which the data live will usually reveal important information about the data. For example, these data live on a 2-dimensional manifold whose axes naturally correspond to the two important invariant transforms implicitly encoded in our examples: rotation and scaling. Building neural models of these transforms provides a convenient and powerful way to learn these representations and naturally facilitates transfer. In addition, it suggests the meta-problem of deciding, given a set of learned transforms, which will be useful (in the complete system insight is the solution to this problem).

The development of creative problem solving in sub-symbolic systems will require innovative research in task similarity metrics, knowledge representation, meta-learning and knowledge transfer mechanisms and will result in a well-grounded (sub-symbolic) computational explanation for several aspects of creativity: *analogy*, *re-representation* and *insight*. In this work, we will analyze and empirically validate an approach to the re-representation problem on artificial and real-world data, showing that we can discover and learn useful transforms in a subsymbolic form that will facilitate analogy and insight in a complete system.

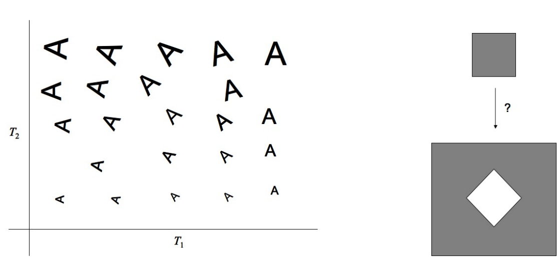

Figure 1: *2-D manifold showing two high-level concept transforms.*