The Consolidation of Task Knowledge
for Lifelong Machine Learning

Daniel L. Silver
Jodrey School of Computer Science
Acadia University
Wolfville, Nova Scotia, Canada B4P 2R6

Introduction

Lifelong Machine Learning, or LML, is concerned with the persistent and cumulative nature of learning (Thrun 1996). LML considers situations in which a learner faces a series of tasks and develops methods of retaining and using prior knowledge to improve the effectiveness (more accurate hypotheses) and efficiency (shorter training times) of learning.

An LML system requires a method of using prior knowledge to learn models for new tasks as efficiently and effectively as possible, and a method of retaining task knowledge after it has been learned. Knowledge retention is necessary for a lifelong learning system, however, it is not sufficient. In (Silver and Poirier 2004) we propose that domain knowledge must be integrated for the purposes of efficient and effective retention and for more efficient and effective transfer during future learning. The process of integration we define as *consolidation*. The challenge for a lifelong learning system is consolidating the knowledge of a new task while retaining and possibly improving knowledge of prior tasks. An interesting aspect of this research is overcoming the *stability-plasticity* problem. The stability-plasticity problem refers to the challenge of adding new information to a system without the loss of prior information (Grossberg 1987).

This paper provides a summary of prior work by the author on the consolidation problem within various lifelong machine learning systems.

Consolidation with MTL and Task Rehearsal

In (Silver and McCracken 2003) a theory of task knowledge consolidation was proposed that uses a large *multiple task learning* or MTL network as the domain knowledge structure and *task rehearsal* as a method of overcoming the stability-plasticity problem. Our early work was heavily influenced by the ideas from (McClelland, McNaughton, and O'Reilly 1994). They suggest a method by which the neocortex of the mammalian brain consolidates new knowledge without loss of previous knowledge. Consolidation occurs through a slow process of interleaved learning of a new and old knowledge within a long-term memory structure of sparse representation. MTL networks are a good choice for consolidated domain knowledge. They have the ability to integrate knowledge from several tasks of a domain and to share this knowledge within their internal representations, provided sufficient representation (hidden nodes) is available. Task rehearsal overcomes the problem of catastrophic forgetting of previously learned tasks stored in a MTL network by relearning synthesized examples of those tasks while simultaneously learning a new task.

The results of experiments on a synthetic domain of seven tasks indicated that the method is capable of sequentially consolidating task knowledge within a domain knowledge MTL network in an efficient and effective manner.

Four Requirements for Consolidation

In (Silver and Poirier 2004) we extended our theory of task knowledge consolidation using MTL networks and task rehearsal. The research confirmed that, under the proper conditions, task rehearsal is capable of sequentially consolidating task knowledge within an MTL network without loss of prior task knowledge. In fact, the experiments indicated an increase in the accuracy of hypotheses for prior tasks as related tasks are consolidated into the MTL network.

This work discovered that the proper conditions for consolidation are (1) a sufficiently large amount of training examples for both new and prior tasks, (2) an abundance of internal representation (number of hidden nodes), (3) a small learning rate so as to ensure slow integration of the new task into the existing representation and (4) a method of preventing the network from over-fitting and therefore creating high-magnitude weight representations. The first condition is met by our LML system because training examples can be synthesized through task rehearsal. The second and third conditions are met with the only impact being an increase in computational space and time for training the neural network. The last condition can be satisfied by employing a weight decay term or an early-stopping technique that uses a validation set of data.

The Effect of Curriculum on Consolidation

In (Poirier and Silver 2005) we investigated the effect of curriculum, *ie.* the order in which tasks are learned, on the consolidation of task knowledge. We researched the use of curriculum in education and developed a theory based on two extremes in the organization of knowledge delivery and

acquisition. One extreme is the *rapid feature capture curriculum*; this ordering of tasks forces the network to create all the internal features as soon as possible. This provides all subsequent tasks with the necessary prior knowledge for developing accurate models. Unfortunately, it means that in the early stages of consolidation there is no knowledge transfer and the mean accuracy of the consolidated tasks may be low. The other extreme is the *gradual feature capture curriculum*; this task ordering spreads out the acquisition of the various features over the greatest number of tasks. This curriculum has the advantage that many new tasks can receive beneficial knowledge from previously consolidated tasks. Its disadvantage is the number of tasks that must be learned in order to acquire all the internal features used in the domain.

We ran two experiments using a synthetic domain of tasks for which we know the shared use of features. Our objectives were (1) to examine the effect on consolidated task accuracy under various short curricula where the order of tasks varies by their sharing of internal features and (2) to examine the effect on task accuracy when learning a larger set of tasks under various curricula. The results showed that curriculum matters and that task ordering affected the accuracy of consolidated knowledge, particularly for the first few tasks that are learned. The results also suggest that as more tasks of a domain are encountered and shared features are learned, the mean accuracy of consolidated domain knowledge tends to converge, regardless of the curriculum.

Continued Practice and Consolidation of Tasks

Our previous research concentrated on sequentially learning each task fully before going on to the next. This approach may be unrealistic, since a lifelong learner may not have access to all the training data for a task at one time. More typically, training data becomes available over the life of the learner. In (O'Quinn, Silver, and Poirier 2005) we investigated the feasibility of learning a task through *continued practice* with small, impoverished data sets as compared to learning the task in a single session with a larger, richer data set. We were interested in the conditions under which the practice of one or more tasks leads to accurate long-term consolidated hypotheses. We proposed that MTL with task rehearsal provides a method of accomplishing this without needing to explicitly store past training examples.

Experimental results showed that even in the presence of unrelated tasks, positive knowledge transfer does occur from past practice sessions to promote an overall improvement in performance on the task. The experiments also revealed that more training examples are required to develop accurate hypotheses through continued practice than would be required if all the examples could be used at once.

Consolidation with csMTL Networks

In (Fowler and Silver 2011) we addressed the problem of knowledge consolidation and the stability-plasticity problem, within a machine lifelong-learning system using a modified MTL neural network, called a *context sensitive* MTL, or *cs*MTL network. A *cs*MTL network is a feed-forward network architecture of input, hidden and output nodes that uses the back-propagation of error training algorithm. The

*cs*MTL network requires only one output node for learning multiple concept tasks. The input layer can be divided into two parts: a set of *primary* input variables for the tasks and a set of inputs that provide the network with the *context* of each training example. The context inputs can be used to associate each training example with a particular task. This research extends the work presented in (Silver, Poirier, and Currie 2008), in which a *cs*MTL network is demonstrated as an effective method of transfer learning. A *cs*MTL-based system can focus on learning shared representation, more than an MTL-based system, because all weight values in the *cs*MTL network are shared between all tasks. The system is meant to satisfy a number of LML requirements including the effective consolidation of task knowledge into a long-term network using task rehearsal, the accumulation of task knowledge from practice sessions, and effective and efficient inductive transfer during new learning.

The experiments demonstrated that consolidation of new task knowledge within a *cs*MTL without loss of prior task knowledge is possible but not consistent for all tasks of the test domain. Our empirical findings indicated that representational transfer of knowledge through the use of prior weights, in addition to functional transfer through task rehearsal, improves retention of prior task knowledge, but at the cost of less accurate models for newly consolidated tasks. We conclude that the stability-plasticity problem is not resolved by our current *cs*MTL-based LML system.

Recent Work and Future Directions

Our most recent efforts have focused on ways to better the results of (Fowler and Silver 2011) by improving the method of task rehearsal of prior tasks. We have developed methods of retaining the probability distribution of examples over the input space for each new task and have used this meta-knowledge to generate more appropriate rehearsal examples. We plan to continue our research into LML systems that use MTL and task rehearsal for knowledge consolidation, however recent results with deep learning architectures suggests that unsupervised learning methods that develop features of the input space need to be incorporated into our approach.

References

Fowler, B., and Silver, D. L. 2011. Consolidation using context-sensitive multiple task learning. In Lingras, P., ed., *Advances in Artificial Intelligence, 24th Conference of the Canadian Artificial Intelligence Association (AI 2011)*, volume LNAI 6657, 128–139. Springer-Verlag.

Grossberg, S. 1987. Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science* 11:23–64.

McClelland, J. L.; McNaughton, B. L.; and O'Reilly, R. C. 1994. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Technical Report PDP.CNS.94.1*.

O'Quinn, R.; Silver, D. L.; and Poirier, R. 2005. Continued practice and consolidation of a learning task. In *Proceedings*

of the Meta-Learning Workshop, 22nd International Conference on Machine Learning (ICML 2005).

Poirier, R., and Silver, D. L. 2005. Effect of curriculum on the consolidation of neural network task knowledge. *Proc. of IEEE International Joint Conf. on Neural Networks (IJCNN 05)*.

Silver, D. L., and McCracken, P. 2003. The consolidation of neural network task knowledge. In Wani, M. A., ed., *Proceedings of the International Conference on Machine Learning and Applications (ICMLA'03)*, 185–192.

Silver, D. L., and Poirier, R. 2004. Sequential consolidation of learned task knowledge. *Lecture Notes in AI, Canadian AI'2004* 217–232.

Silver, D. L.; Poirier, R.; and Currie, D. 2008. Inductive tranfser with context-sensitive neural networks. *Machine Learning* 73(3):313–336.

Thrun, S. 1996. Is learning the n-th thing any easier than learning the first? In *Advances in Neural Information Processing Systems*, 640–646. The MIT Press.