

A Measure of Relatedness for Selecting Consolidated Task Knowledge

Daniel L. Silver and Richard Alisch

Jodrey School of Computer Science

Acadia University

Wolfville, Nova Scotia, Canada B4P 2R6

email: danny.silver@acadiau.ca

Abstract

The selective transfer of task knowledge is studied within the context of multiple task learning (MTL) neural networks. Given a consolidated MTL network of previously learned tasks and a new primary task, T_0 , a measure of task relatedness is derived. The existing consolidated MTL network representation is fixed and an output for task T_0 is connected to the hidden nodes of the network and trained. The cosine similarity between the hidden to output weight vectors for T_0 and the weight vectors for each of the previously learned tasks is used as measure of task relatedness. The most related tasks are then used to learn T_0 within a new MTL network using the task rehearsal method. Results of an empirical study on two synthetic domains of invariant concept tasks demonstrate the method's ability to selectively transfer knowledge from the most related tasks so as to develop hypotheses with superior generalization.

Introduction

Many problems to which machine learning is applied suffer from a deficiency of training examples. Learning theory tells us that the development of a sufficiently accurate model from a practical number of examples depends upon an appropriate *inductive bias*, one source being *prior task knowledge* (Mitchell 1980). Most machine learning systems do not use knowledge from potentially related tasks when inducing a model of a classification task from a set of supervised training examples, instead they focus on a single task learning approach. Lacking a theory of selective knowledge transfer (Thrun 1997; Caruana 1997), we are working to develop one and to apply it in practical problems of sequential lifelong learning, such as developing more accurate medical diagnostic models from a small number of patients by using previously learned models. η MTL, a modified version of the multiple task learning (MTL) method of parallel functional transfer is introduced in (Silver & Mercer 2001). An η MTL artificial neural network (ANN) biases the induction of a hypothesis for a new primary task based on the relatedness of unconsolidated secondary tasks to the new task. Various functional and structural measures of relatedness were

proposed and compared. The conclusion was that a consolidated representation of all previously learned tasks would provide the best source for knowledge transfer because it provides the basis for measuring deep structural similarity between tasks. In (Silver & Poirier 2004) it is demonstrated that new tasks can be consolidated within an existing MTL network with insignificant loss to the generalization accuracy of previously learned tasks. In fact, minor increases in the generalization of related prior tasks were observed. Consolidation is accomplished through the rehearsal of previously learned tasks while slowly integrating a new task. This paper follows up these works by investigating a measure of task relatedness based on a consolidated MTL network of prior task knowledge. The measure is applied to a challenging set of two task domains where each domain shares an invariant in the input space.

Background

The constraint on a learning system's hypothesis space, beyond the criterion of consistency with the training examples, is called inductive bias (Mitchell 1980). For example, Occam's Razor suggests a bias for simple over more complex hypotheses. Inductive bias is essential for the development of an hypothesis with good generalization from a practical number of examples. Ideally, a life-long learning system can select its inductive bias to tailor the preference for hypotheses according to the task being learned. One type of inductive bias is prior knowledge of the task domain. The retention and use of task domain knowledge as a source of inductive bias remains an open problem in machine learning (Thrun 1997; Caruana 1997).

In (Silver & Mercer 2001) *knowledge-based inductive learning* is defined as a life-long learning method that uses knowledge of the task domain as a source of inductive bias. As with a standard inductive learner, training examples are used to develop an hypothesis of a classification task. However, unlike a standard learning system, knowledge from each hypothesis is saved in a long-term memory structure called domain knowledge. When learning a new task, aspects of domain knowledge are selected to provide a positive inductive bias to the learning system. The result is a more accurate hypothesis developed in a shorter period of time. The method relies on the transfer of knowledge from one or more prior secondary tasks, stored in domain knowledge, to

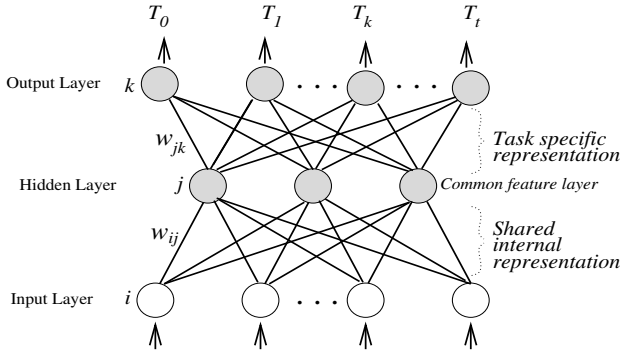


Figure 1: A multiple task learning (MTL) network with an output node for each task being learned in parallel.

the hypothesis of the new task. Consequently, the problem of selecting an appropriate bias becomes one of selecting the most related task knowledge for transfer.

Knowledge Transfer in MTL Networks

Multiple task learning (MTL) neural networks are one of the better documented methods of knowledge transfer (Caruana 1997). An MTL network is a feed-forward multi-layer network with an output for each task that is to be learned. The standard back-propagation of error learning algorithm is used to train all tasks in parallel. Consequently, MTL training examples are composed of a set of input attributes and a target output for each task. Figure 1 shows a simple MTL network containing a hidden layer of nodes that are common to all tasks. The sharing of internal representation is the method by which inductive bias occurs within an MTL network (Baxter 1996). MTL allows two or more tasks to share all or part of the common feature layer to the extent to which it is mutually beneficial. The more that tasks are related the more they will share representation and create positive inductive bias.

Much of our research has focused on methods of selective knowledge transfer that distinguishes between knowledge from related and unrelated tasks. In (Silver & Mercer 2001), η MTL, a modified version of MTL, is created to provide a solution to the problem of selective transfer of task knowledge. η MTL will be explained in the section on task relatedness.

Sequential Learning through Task Rehearsal

In (Silver & Mercer 2002) the *task rehearsal method* was introduced as a knowledge-based inductive learning system that is able to retain and recall unconsolidated task knowledge. Building on the theory of pseudo-rehearsal (Robins 1995), previously learned but unconsolidated task representations are used to generate *virtual examples* as a source of functional knowledge. After a task T_k has been successfully learned (to a specified level of generalization error), its hypothesis representation is saved in domain knowledge. This representation acts as a surrogate for the space of input-output examples that defines task T_k . Virtual examples of the input-output space for T_k can be produced (with the

same level of generalization error) by passing inputs to the domain knowledge representation for T_k and recording the outputs. When learning a new task, T_0 , the domain knowledge representations for tasks $T_1 \dots T_k \dots T_t$ are used to generate corresponding *virtual* output values from the set of T_0 training examples. The resulting set of virtual examples is used to relearn or *rehearse* the domain knowledge tasks in parallel with the learning of T_0 in an MTL or η MTL network. It is through the rehearsal of previously learned tasks that knowledge is transferred to the new task.

A method of sequential consolidation of task knowledge using an MTL network and Task Rehearsal is proposed and tested in (Silver & Poirier 2004). This method overcomes the stability-plasticity problem originally posed by (Grossberg 1987) taken to the level of learning sets of tasks as opposed to learning sets of examples. The stability-plasticity problem is the difficulty of learning a new example within a neural network while trying to maintain knowledge of previously learned examples. The paper shows that the consolidation of new task knowledge without loss of existing task knowledge is possible given sufficient number of training examples, sufficient internal representation for all tasks, slow training using a small learning rate and a method of early stopping to prevent over-fitting and therefore the growth of high magnitude weights. The benefits of retaining domain knowledge in a consolidated form are (1) more efficient and effective storage of individual hypotheses and (2) the basis by which to determine the structural similarity between hypothesis representations. The following section proposes one method of measuring the relatedness between a new primary task and the tasks knowledge retained with a consolidated MTL network.

Task Relatedness and Selective Transfer

The above background material has led to a theory of selective knowledge transfer from an MTL network of consolidated domain knowledge. In this section we summarize those aspects of the theory that are concerned with (1) a framework for adjusting inductive bias within an MTL network via a measure of task relatedness (2) the nature of task relatedness and (3) a method of using consolidated MTL domain knowledge to measure task relatedness based on structural similarity.

Framework for a Measure of Task Relatedness

Consider an objective function to be minimized by the BP algorithm across all task outputs of a MTL network that focuses on the development of the best hypothesis for the primary task. Let $\hat{E} = E_0 + \sum_k E_k$, where E_k is the error on the training examples for the secondary tasks and E_0 is the error on the primary task training examples. By gradient descent the appropriate change to a weight w_{jk} at an output node k is given by $\Delta w_{jk} = -\eta \frac{\partial \hat{E}}{\partial E_k} \frac{\partial E_k}{\partial w_{jk}}$ where η is the learning rate. Under these conditions, $\frac{\partial \hat{E}}{\partial E_k}$, the rate of change of the overall error with respect to the rate of change of the error for task k , can be considered the weight of importance of task k for learning the primary task. We define

this weight of importance to be the *measure of relatedness*, R_k , between the primary task and each of the secondary tasks; that is $\Delta w_{jk} = -\eta R_k \frac{\partial E_k}{\partial w_{jk}}$. Thus, an appropriate measure of relatedness, R_k , for a secondary task, T_k , must regulate the impact of the task error, E_k , on the formation of shared internal representation. This modified version of the standard back-propagation learning algorithm for MTL is called the η MTL algorithm because R_k tailors the learning rate η for each secondary task (Silver & Mercer 2002).

Let $R_0 = 1$ for the primary task, T_0 , and let $0 \leq R_k \leq 1$ for all other tasks, T_k , thereby constraining the learning rate for any secondary task to be at most η . If $R_k = 1$ for all k , we have standard MTL. Alternatively, if $R_0 = 1$ and $R_k = 0$ for all k , we have standard single task learning (STL) of the primary task. In this way, the η MTL framework generalizes over STL and MTL.

Nature of Task Relatedness

Critical to the transfer of knowledge from a pool of secondary tasks to a primary task is some measure of relatedness between those tasks (Thrun 1997; Caruana 1997). A brute force approach to determining the relatedness between the primary task and each secondary tasks would be to have the learning system learn T_0 in parallel with every other secondary task and record the effectiveness of each T_0 hypothesis. However, if combinations of secondary tasks are considered then the method would be impractical because of the factorial growth in time complexity. The next section discusses two categories of *a priori* measures of relatedness that attempt to overcome this problem.

Measures of Relatedness

Measures of relatedness can be grouped into at two broad categories (Silver & Mercer 2001); (1) *surface similarity* measures or (2) *structural similarity* measures. Surface similarity measures are estimates based on shallow, easily perceived, external similarity which is a measure of the external functional similarity. Surface similarity measures are most typically based on mathematical relationships between the output values of the primary task and each secondary task; for example the *coefficient of linear correlation*, r , between the target values of T_k and T_0 . We have previously used a measure of relatedness $R_k = |r|$ with some success (Silver & Mercer 2001).

The current paper is interested in judging relatedness between pattern recognition tasks where each task must learn an invariant of the concept pattern in the input space. Unfortunately, surface similarity measures suffer from an inability to capture such complex invariant relations between tasks.

Structural similarity measures are estimates of relatedness based on deep, often complex, internal feature similarity such as the degree to which two MTL hypotheses utilize shared internal representation to produce accurate approximations to tasks. Cognitive researchers believe that knowledge transfer begins by using surface similarity as the preliminary search key into domain knowledge (Robins 1995). Structural similarity is subsequently used to scrutinize the selected tasks at a finer level of detail — at a level where

the internal “hidden” representation of the tasks are compared. We propose that structural measures provide an opportunity to capture the measure of relatedness between invariant tasks.

Structural similarity measures can be based on the similar use of the common feature layer by each task specific portion of an MTL network, or the similarity of the weight space representations in the task specific portions of an MTL network. Consider the η MTL network shown in Figure 1. Each secondary hypotheses h_k differs from the primary hypothesis h_0 only in terms of their common feature layer node to output node weight vectors, W_0 and W_k . Co-linearity of W_0 and W_k is a measure of the relatedness between h_0 and h_k . Let θ be the angle between W_k and W_0 . A measure of structural similarity is given by $R_k = |\cos \theta|$.

Consolidated MTL Networks & Task Relatedness

Consider a consolidated MTL 3-layer network (as per Figure 1) with a set of previously learned tasks such as T_k from a domain and training data for a new primary task T_0 . We propose to measure the relatedness of the new task to the prior tasks in the following way. First, fix the representation of the exiting MTL network including the hidden to output weights, W_k , for each task. Second, add an output node for T_0 to the hidden nodes and set its vector of connection weights, W_0 , to small random values. Third, train the network to learn T_0 as accurately as possible using cross-validation. Finally, determine the mean absolute cosine of the angle, $|\cos \theta|$, between the weight vectors W_0 and each W_k . Set R_k equal to this cosine similarity estimate. The preliminary training of the primary task hypothesis does not take long because the only weights that changed are in W_0 .

Two methods are proposed for selecting related tasks based on R_k . The first method called “0.5 cut-off” selects all tasks within the consolidated network with an $R_k > 0.5$. The second method called “average R_k cut-off” calculates the mean R_k across all tasks within the consolidated network, \bar{R}_k . It then selects all tasks with an $R_k > \bar{R}_k$.

One significant challenge for this method concerns the accuracy of the preliminary hypothesis for T_0 . Clearly, the validity of the weights depends upon the generalization accuracy of the hypothesis. If a sufficiently accurate hypothesis cannot be developed then it is reasonable to conclude that there are no related tasks within the consolidated domain knowledge MTL network. Consequently, there is no value in using the measures of relatedness to selectively transfer prior knowledge. For the purposes of this paper, the R_k value is accepted only if the preliminary hypotheses for T_0 are learned with a sufficient level of mean accuracy. The issue of preliminary training accuracy will be investigated further in future work.

Empirical Studies

To test the hypothesis of selective functional transfer using a consolidated MTL network and the weight-based structural measure of task relatedness, a series of experiments were conducted using tasks from two distinct domains. Both domains share the same inputs and contain nonlinearly separable tasks that identify invariant concepts in the input space.

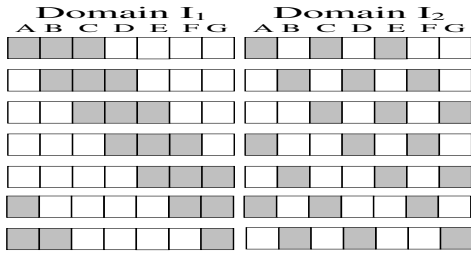


Figure 2: The two Invariant domains. Each task must recognize a pattern of three attributes that are invariant over the input space in two different ways.

These tasks were chosen because they are not amenable to simple surface measures of relatedness like linear correlation.

The Invariant Domains

Figure 2 characterizes the tasks from the two domains. Both domains exhibit translational invariance in the shared input space; for each task a pattern of three input values is invariant over the seven attributes. Each input attribute can have three possible values: 0.5 meaning background (B), < 0.5 meaning a low (L) value, and > 0.5 meaning a high (H) value. Each task in each domain identifies one of five patterns in the inputs: the primary task, T_0 , identifies the pattern HLH and the secondary tasks identify each of the following patterns: T_1 , HHH; T_2 , LHL; T_3 , LLL and T_4 , LHH.

The difference between the domains is the mapping of the three invariant attribute values over the seven inputs. Invariant domain one, I_1 , contains tasks where the pattern is mapped to any three contiguous inputs (with wrap-around); for example BLHHBBB. Invariant domain two, I_2 , contains tasks with the pattern mapped to five contiguous inputs with a background input between each attribute (with wrap-around); for example BLBHBHB. Labelling the inputs as A through G, the five tasks of each domain can be expressed mathematically, for example task I_1T_0 would be: $(A > 0.5 \wedge B < 0.5 \wedge C > 0.5) \vee (B > 0.5 \wedge C < 0.5 \wedge D > 0.5) \vee \dots \vee (G > 0.5 \wedge A < 0.5 \wedge B > 0.5)$. This problem is equivalent to learning to recognize a 3-pixel pattern in a data stream that is seven pixels wide, independent of the pattern's starting position.

The examples were generated as follows: all inputs were initialized to the background value, three random attribute values were generated and mapped to a random starting input and finally the target output (1 if true and 0 if false) was generated for each task according to the task's attribute pattern. The examples for each task were evenly split across all seven possible mappings of the patterns to the inputs.

The tasks of the two domains should compete for internal representation at the common feature layer of the consolidated MTL network. We expect that I_1 tasks will tend to share a subset of internal representation (hidden nodes) different from that of the I_2 tasks that captures the I_1 invariance mapping. Ultimately, those secondary tasks that contribute most positively to the development of an accurate hypoth-

esis for T_0 will be considered the most related. A surface similarity measure such as correlation does poorly on judging the relatedness between any tasks *a priori*. For example the $|r|$ values between I_1T_0 and each of the other eight tasks $I_1T_1, \dots, I_1T_4, I_2T_1, \dots, I_2T_4$ are respectively 0.1137, 0.1200, 0.1078, 0.1156, 0.1137, 0.1200, 0.1078, 0.1156.

Experiments

The following experiments provide evidence that a structural measure of relatedness based on consolidated domain knowledge can be used to promote a positive transfer of knowledge for the development of a more effective primary hypotheses. The first experiment examines the learning the primary task, T_0 , under STL and η MTL using all tasks of both domains, all tasks of each domain, and each individual task. This provides a base line for learning T_0 and the relatedness of T_0 to each of the secondary tasks. The second experiment demonstrates (1) the calculation of R_k based on the representation of the consolidated MTL network containing the secondary tasks of both domains and (2) the benefit of using R_k for learning T_0 under η MTL.

The consolidated MTL hypotheses for all secondary tasks were developed using 500 training and 500 validation examples and then tested against the 1000 test examples. The hypotheses were shown to have accuracies no less than 90%. The preliminary training and final training of the primary task I_1T_0 uses an impoverished training set of only 100 training examples. This is accomplished by marking the target value for 400 of the primary task examples as unknown. In our software target values marked as *unknown* make zero contribution to weight modifications for that task.

The neural networks used in the experiments have an input layer of 7 nodes, one hidden layer (common feature layer) of 14 nodes, and an output layer of 9 nodes, one for each secondary task and I_1T_0 . Each node uses a standard sigmoid activation function. The number of hidden nodes is sufficient for all tasks. One hidden node is required to act as a receptive field for each of the seven possible attribute mappings for each domain. Therefore, 14 hidden nodes provide sufficient internal representation for the two domains of tasks under MTL (Caruana 1997). In both experiments the mean squared error cost function is minimized by the BP algorithm. The base learning rate, η , is 0.005 and the momentum term is 0.9. Random initial weight values are selected in the range -0.1 to 0.1 for all runs.

One trial consists of 5 repetitions using different random initial weight vectors. To reduce experimental variance, the same set of initial weights are used by the various learning methods during one repetition. Performance of the methods is based on the generalization accuracy of the hypotheses for the primary task developed from the impoverished training set. Generalization accuracy is the mean proportion of correct classifications made by the primary hypotheses against an independent test set over the repetitions. Classification is subject to a cut-off value of 0.5 (any example with a value ≥ 0.5 is considered class 1). A difference of means hypothesis test (1-tailed, paired) based on a t -distribution will determine the significance of the difference between the mean accuracy for any two trials.

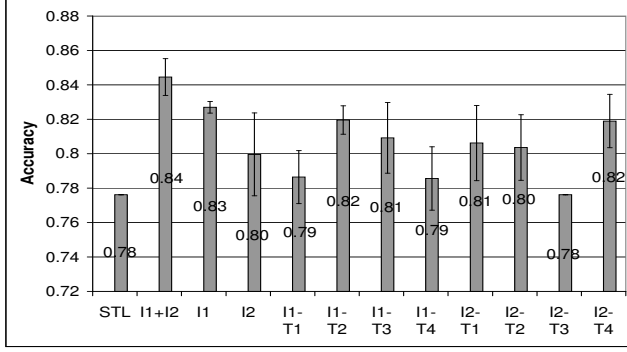


Figure 3: Experiment 1: Accuracy of T_0 hypotheses under inductive bias provided by various secondary tasks.

Experiment 1: Inductive bias from various tasks. This experiment examines the inductive bias provided to the primary task, T_0 , within an η MTL network when learned in parallel with a variety of secondary tasks. The following parameter settings are tested during 12 trials: the primary task, T_0 , with no secondary tasks (all $R_k = 0.0$); all 8 secondary tasks of both domains, I_1 and I_2 (all $R_k = 1.0$); the 4 secondary tasks of domain I_1 ; the 4 secondary tasks of domain I_2 ; and 8 trials of η MTL where only one of the secondary tasks is learned in parallel with T_0 ($R_0 = 1.0$, $R_k = 1.0$, $R_i = 0.0$, $i \neq k$ for trial k).

Results and Discussion. Figure 3 presents the performance results for hypotheses developed by each of the learning methods over 5 repetitions. Shown are the mean accuracy (and 95% confidence interval) on a 1000 example test set. The hypotheses for the primary task developed from the impoverished training set (100 training examples) using STL performed the worst on the test set. Parallel learning of all secondary tasks from domains I_1 and I_2 under η MTL developed the best hypotheses with an accuracy of 85%. As expected the most helpful domain of tasks were those of I_1 with a statistical difference of 3% ($p=0.091$) over I_2 .

When $R_0 = 1.0$ and only one other $R_k = 1.0$, the η MTL method isolates the inductive bias to task T_k . Figure 3 shows that the most consistently effective T_0 hypotheses were those developed in parallel with tasks T_2 , T_3 from domain I_1 and surprisingly T_1 , T_2 and T_4 from domain I_2 . The positive inductive bias from the domain I_2 tasks was not expected nor is it in agreement with the results from MTL learning with all tasks of each domain. We conclude that the mixture of tasks being learned in parallel has a significant impact on the inductive bias from each task. In additional, we conclude that there are similarities between the tasks across the domains in spite of our attempt to create difference in the invariants.

Experiment 2: Selective Inductive Bias Using R_k . This experiment demonstrates the calculation of the R_k values based on the use of an existing MTL network that contains

Task	$R_k = \cos \theta $	0.5 cut-off	ave. R_k cut-off
I_1T_1	0.158		
I_1T_2	0.430		selected
I_1T_3	0.585	selected	selected
I_1T_4	0.227		
I_2T_1	0.133		
I_2T_2	0.181		
I_2T_3	0.141		
I_2T_4	0.604	selected	selected

Table 1: Measure of relatedness values and the selection of most related tasks.

accurate representations of all I_1 and I_2 secondary tasks. It then examines the inductive bias provided to the primary task, T_0 , within an η MTL network when the R_k values are used to select the most related tasks for knowledge transfer. Both the “0.5 cut-off” and “average cut-off” methods of using R_k to select the most related tasks are tested.

A 7-14-8 MTL network was used to develop accurate hypotheses for the eight secondary tasks T_1, \dots, T_4 of the two domains. As per our theory, the representation of the existing MTL network including the hidden to output weights, W_k , was then fixed. An output node for T_0 was added by connecting it to the hidden nodes and setting its weights, W_0 , to small random values. The network was trained five different times to learn T_0 using an impoverished training set (100 examples). The mean absolute cosine of the angle, $|\cos \theta|$, between the weight vectors W_0 and each W_k was calculated. This mean value is used as the value for each R_k . The R_k values were deemed acceptable if the T_0 task trained to a level of 70% accuracy or better. The final selections of most related tasks were based on the “0.5 cut-off” and the “average R_k cut-off” method as described earlier.

The final training of the primary task, T_0 , uses the same network and parameter configuration as Experiment 1; however, all connection weights are initialized to small random values. The R_k values for the most related tasks were set to 1.0; all other R_k values were set to 0.0. This ensures that only the related tasks will be learned in parallel with T_0 .

Results and Discussion. Table 1 shows the R_k values generate by the cosine similarity using the consolidated MTL network weights. Observe the similarity of these values to that of the inductive bias provided to the T_0 hypotheses by each of the domain knowledge tasks shown in Figure 3. The linear correlation, r^2 , of these values is 0.712, indicating a good deal of similarity. Only the R_k values for I_2T_1 and I_2T_2 deviate significantly from the inductive bias results. This provides us with some confidence that the proposed measure of task relatedness has merit.

Figure 4 presents the performance results for T_0 hypotheses developed by each of the learning methods using impoverished training sets over 5 repetitions. Shown are the mean accuracy (and 95% confidence interval) on a 1000 example test set. The T_0 hypotheses developed using the measure of relatedness methods perform as well as or better than all

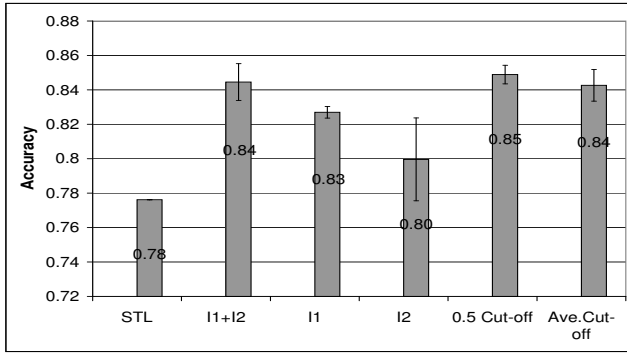


Figure 4: Experiment 2: Accuracy of T_0 hypotheses under inductive bias provided by selective transfer of secondary task knowledge.

other hypotheses. This includes those developed under the various choices of tasks or domains of Figure 3. In particular, the hypotheses developed under the 0.5 cut-off method that chose I_1T_3 and I_2T_4 as the most related tasks is statistically better ($p < 0.092$) than all variations except when T_0 is learned in parallel with all tasks ($I_1 + I_2$).

Summary and Conclusion

This paper has introduced a method of measuring task relatedness between a new primary task, T_0 and previously learned tasks consolidated within an MTL neural network. The measure is based on the structural similarity between the representation of each existing task T_k and a preliminary hypothesis developed for T_0 from the consolidated MTL network. The resulting measures, R_k , are used to select those secondary tasks that are most related to T_0 for learning within a new MTL network. We use an η MTL network for learning T_0 because it provides a framework for controlling the selection of secondary tasks and therefore inductive bias through knowledge transfer.

The results of Experiment 1 indicate that a significant transfer of knowledge occurs between secondary tasks from the two invariant domains of tasks to the primary task. It further indicates that in composite, tasks from the same domain I_1 provide the strongest positive inductive bias to T_0 . However, the experiments also show that certain tasks from invariant domain I_2 positively bias the development of a hypothesis for T_0 . In fact they do so more individually than they do in composite. We conclude that (1) the mixture of tasks being learned in parallel has a significant impact on the transfer of knowledge from each task and (2) that there are similarities between the tasks that were not anticipated. This indicates that further work is required on the nature of relatedness between tasks within a consolidated representation.

The results of Experiment 2 show that a structural measure of task relatedness based on consolidated domain knowledge can be used to selectively transfer knowledge and adjust inductive bias. The measures of relatedness returned

by the method correlated very strongly with the performance of the T_0 hypotheses when learned with each of the individual secondary tasks ($r^2 = 0.712$). This suggests that the cosine similarity measure captures aspects of the sharing of internal representation between T_0 and secondary task hypotheses within the consolidated MTL network.

The “0.5 cut-off” method produces hypotheses with the highest mean accuracy over all other variations. The method consistently chose I_1T_3 and I_2T_4 as the most related tasks. However, the selective transfer results are not significantly better than when all tasks are learned at the same time in one MTL network. Our goal in future research is to ensure greater diversity between the two task domains.

Will the measure of relatedness scale up? Theoretically, the method of measuring task relatedness should scale to large numbers of secondary tasks within a consolidated MTL network. Adjustment of an appropriate cut-off will ensure only the most related tasks are selected for learning within an η MTL network with the new task. The larger problem is that of sequential consolidation of new task knowledge into an ever increasing consolidated MTL network. We have previously reported advances in this area (Silver & Poirier 2004) and are currently investigating the effect of curriculum and practise on consolidation.

Acknowledgement: This research has been funded in part by the Government of Canada through NSERC.

References

- Baxter, J. 1996. Learning model bias. In Touretzky, D. S.; Mozer, M. C.; and Hasselmo, M. E., eds., *Advances in Neural Information Processing Systems*, volume 8, 169–175.
- Caruana, R. A. 1997. Multitask learning. *Machine Learning* 28:41–75.
- Grossberg, S. 1987. Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science* 11:23–64.
- Mitchell, T. M. 1980. The need for biases in learning generalizations. *Readings in Machine Learning* 184–191. ed. Jude W. Shavlik and Thomas G. Dietterich.
- Robins, A. V. 1995. Catastrophic forgetting, rehearsal, and pseudorehearsal. *Connection Science* 7:123–146.
- Silver, D. L., and Mercer, R. E. 2001. Selective functional transfer: Inductive bias from related tasks. *Proceedings of the International Conference on Artificial Intelligence and Soft Computing (ASC2001)* 182–189.
- Silver, D. L., and Mercer, R. E. 2002. The task rehearsal method of life-long learning: Overcoming impoverished data. *Advances in Artificial Intelligence, 15th Conference of the Canadian Society for Computational Studies of Intelligence (AI’2002)* 90–101.
- Silver, D. L., and Poirier, R. 2004. Sequential consolidation of learned task knowledge. *Lecture Notes in Artificial Intelligence, 17th Conference of the Canadian Society for Computational Studies of Intelligence (AI’2004)* 217–232.
- Thrun, S. 1997. Lifelong learning algorithms. *Learning to Learn* 181–209.1 Kluwer Academic Publisher.