Transfering Nonlinear Representations using Gaussian Processes with a Shared Latent Space

Raquel Urtasun[1], Ariadna Quattoni[1], Neil D. Lawrence[2] and Trevor Darrell[1]

[1] Massachusetts Institute of Technology, Cambridge, MA 02139 USA
[2] School of Computer Science, University of Manchester, M13 9PL, U.K.

When faced with a new task, it is advantageous to exploit knowledge and structures found useful in solving related problems. A common paradigm to exploit such knowledge is to learn a feature space from previous tasks and transfer that representation to a future task (Baxter, 1997; Caruana, 1997; Thrun, 1996). Ideally, the transferred representation is of lower dimension than the raw feature space, and the set of functions implied by the new representation still contains the optimal classifier for the new task. When this is the case, the new task can be learned more robustly and/or with fewer training examples in the transferred space than in the raw space (Ando & Zhang, 2005).

In this paper we propose a novel approach to transfer learning based on discovering a low-dimensional, non-linear latent space jointly across tasks in a Gaussian Process framework, and transferring that space to future tasks.

Transfer of probabilistic representations has been explored in a Gaussian Processes (GP) paradigm, by explicitly sharing a covariance function and/or kernel hyperparameters across tasks (Lawrence & Platt, 2004; Yu et al., 2005). However, the amount of sharing that this introduces is somehow limited.

More recently, Bonilla et al. extended previous approaches to model the relateness between tasks with a parametric (Bonilla et al., 2007a) and non-parametric covariance (Bonilla et al., 2007b). However, it is often the case that the relateness is not task-dependent but sample dependent. In other words some samples of task A might be related to task B, while some others samples might not be related at all. Our method estimates the relateness of the different samples by estimating a low dimensional representation that is shared across tasks. Samples that are related are close in latent space.

Probably the closest approach to ours is the one developped by Teh et al. (Teh et al., 2005), that proposed a semiparametric factor model that models the dependencies by a set of Gaussian Processes linearly mixed. In contrast, our method learns directly a non-linear low dimensional latent space that is shared accross tasks. When using a Gaussian noise model, our model is computationaly less expensive since it does not require a variational approximation.

Experiments on digit recognition tasks indicate that the ability to transfer non-linear, low-dimensional features across problems can provide significant performance improvements, especially when the target task has relatively few training examples compared to the source problems used to learn the latent space. Baseline experiments confirm that learning the shared latent space discriminatively is important; Semi-supervised learning underperformed transfer learning with representations learned discriminatively.

Bibliography

Ando, R. K., & Zhang, T. (2005). A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, *6*, 817–8535.

Baxter, J. (1997). A bayesian/information theoretic model of learning to learn via multiple task sampling. *Machine Learning*, *28*.

Bonilla, E. V., Agakov, F. V., & Williams, C. K. I. (2007a). Kernel Multi-task Learning using Task-specific Features. .

Bonilla, E. V., Chai, K. M. A., & Williams, C. K. I. (2007b). Multi-task gaussian process prediction. *Neural Information Processing Systems*. MIT Press.

Caruana, R. (1997). Multitask learning. *Machine Learning*, *28*, 41–75.

Lawrence, N. D., & Platt, J. C. (2004). Learning to learn with the informative vector machine. *International Conference in Machine Learning*.

Teh, Y. W., Seeger, M., & Jordan, M. I. (2005). Semiparametric latent factor models.

Thrun, S. (1996). Is learning the n-th thing any easier than learning the first? *Neural Information Processing Systems*.

Yu, K., Tresp, V., & Schwaighofer, A. (2005). Learning Gaussian processes from multiple tasks. *International Conference in Machine Learning* (pp. 1012–1019).