Available Abstarcts

Tuesday, July 11 (Room 1.03, Malet Place, New Engineering Building)

10:00–10:50: *Shai Ben-David* (University of Waterloo, Canada)
Inductive transfer via embeddings into a common feature space

We consider the situation in which there is a basic learning task but different sub-tasks define different data generating distributions. Examples include learning to identify spam for various different email users, or parts-of-speech tagging for different text corpora. Our goal is to allow the use of training data coming from one sub-task for prediction under another sub-task distribution. Our suggested solution is to embed the domain of the different tasks into a feature space that retains sufficient information for label prediction while, at the same time, the distributions induced on that feature space by the of the different tasks are similar to each other. We prove a generalization error bound that is based on these two properties. We apply our paradigm to the parts-of-speech tagging problem. In this application, the feature space we use is based on triangulation of the original unlabeled data set around pivot points that behave similarly under the different sub-task distributions.
This is joint work with John Blitzer, Koby Crammer and Fernando Pereira.

11:00–11:50: *Andreas Argyriou* (University College London, UK)
Multi-task feature learning

We present a method for learning a low-dimensional representation which is shared across a set of multiple related tasks. The method builds upon the well-known 1-norm regularization problem using a new regularizer which controls the number of learned features common for all the tasks. We show that this problem is equivalent to a convex optimization problem and develop an iterative algorithm for solving it. The algorithm has a simple interpretation: it alternately performs a supervised and an unsupervised step, where in the latter step we learn common-across-tasks representations and in the former step we learn task-specific functions using these representations. We report experiments on a simulated and a real data set which demonstrate that the proposed method dramatically improves the performance relative to learning each task independently. Our algorithm can also be used, as a special case, to simply select – not learn – a few common features across the tasks.
This is joint work with Theodoros Evgeniou and Massimiliano Pontil

14:00–14:50: *Tong Zhang* (Yahoo!, USA)
Learning with structured inputs

I will present a novel approach to semi-supervised learning that employs a method which we refer to as structural learning (aka multi-task learning). The idea is to learn predictive structures from many auxiliary problems that are created from the unlabeled data (and are related to the target problem), and then transfer the learned structure to the supervised target problem.

In the first part, I will give a high level description of the general approach, a specific bi-linear structure model we use, and then explain how we generate auxiliary problems that are related to the target task. I will then show some empirical results. For example, this method produces performance higher than the previous best results on some standard NLP benchmark tests. It is also highly effective for some other problems such as image classification and even information retrieval. In particular, I will describe our successful participation in 2005's TREC genomics ad hoc retrieval task using this idea.

In the second part, I will present a general framework for learning structures including its learning theoretical analysis. Under this framework, I will investigate the theoretical justification of the bi-linear structure model used in our experiments, and show that the resulting formulation can be solved by an iterative SVD procedure. The relationship of this framework and transfer learning, and Bayesian hierarchical models will be discussed.

This is a joint work with Rie Ando of IBM TJ Watson Research Center

15:00–15:50: *Sayan Mukherjee* (Duke University, USA)
Estimation of gradients and coordinate covariation in classification

We introduce an algorithm that simultaneously estimates a classification function as well as its gradient in the supervised learning framework. The motivation for the algorithm is to find salient variables and estimate how they covary. An efficient implementation with respect to both memory and time is given. The utility of the algorithm is illustrated on simulated data as well as a gene expression data set. An error analysis is given for the convergence of the estimate of the classification function and its gradient to the true classification function and true gradient.

Wednesday, July 12 (Room 1.03, Malet Place, New Engineering Building)

10:00–10:50: *Andreas Maurer* (Stemmer Imaging, Germany)
Slow subspace learning from stationary processes

The talk presents a method of unsupervised learning from stationary, vector-valued processes. The method selects a subspace on the basis of an objective which can be used to bound the expected classification error for a family of tasks posessing a temporal continuity property. We prove bounds on the objective's estimation error in terms of mixing coefficients and consistency for absolutely regular processes. Experiments with image processing demonstrate the algorithms ability to learn geometrically invariant feature maps."

11:00–11:50: *Tom Heskes* (Radboud University Nijmegen, The Netherlands)
Multitask learning: the Bayesian way

Multi-task learning lends itself particularly well to a Bayesian approach. Cross-inference between tasks can be implemented by sharing parameters in the likelihood model and the prior for the task-specific model parameters. Choosing different priors, one can implement task clustering and task gating. Throughout my presentation, predicting single-copy newspaper sales will serve as a running example.

14:00–14:50: *Claudio Gentile* (Universita' dell'Insubria, Italy)
Top-down vs. bottom-up methods for hierarchical classification

We deal with hierarchical classification in the general case when an instance could be associated with multiple and/or partial paths in a given taxonomy. We approach the problem from different perspectives: "top-down vs. bottom-up", but also "on-line vs. batch" and "theoretical vs. experimental". In this talk, we briefly present our recent research experience on this subject matter.
Joint work with N. Cesa-Bianchi and L. Zaniboni.

15:00–15:50: *Antonio Torralba* (Massachusetts Institute of Technology, USA)
Learning shared representations for object recognition

ABSTRACT TBA

Thursday, July 13 (Room 1.03 Malet Place, New Engineering Building)

10:00–10:50: *Yasemin Altun* (Toyota Technological Institute, USA)
Supervised and semi-supervised learning for complex/structured output prediction

Many real-world classification problems involve the prediction of multiple inter-dependent variables forming some structural dependency. Recent progress in machine learning has mainly focused on supervised classification of such structured variables. I will talk about these approached and then investigate structured classification in a semi-supervised setting. I'll present a discriminative approach that utilizes the intrinsic geometry of input patterns revealed by unlabeled data points and present a maximum-margin formulation of semi- supervised learning for structured variables. Unlike transductive algorithms, this formulation naturally extends to new test points.

11:00–11:50: *Florence d'Alché* (Universite' d'Evry, France)
Output kernel tree

In this paper, we generalize tree-based methods to the prediction in structured output space. The extension is based on a kernelization of the algorithm that allows one to grow trees as soon as a kernel can be defined on the output space. The resulting algorithm, called output kernel trees (OK3), generalizes classification and regression trees in a principled way. It inherits several features of these methods such as interpretability, robustness to irrelevant variables, and input scalability. OK3 can also directly take advantage of most techniques that enhance standard trees, e.g. pruning, ensemble methods, attribute selection and ranking. We illustrate the algorithm on an artificial image reconstruction task. The algorithm still works when the output values are not observed but instead only a Gram matrix is given over the learning sample, in which case, OK3 is able to generalize the underlying kernel to new data. We exploit this feature on a problem of biological network inference.

14:00–14:50: *Juho Rousu* (University of Helsinki, Finland)
Efficient max-margin markov learning via conditional gradient and probabilistic inference

We present a general and efficient optimisation methodology for for max-margin sructured classification tasks. The efficiency of the method relies on the interplay of several techiques: marginalization of the dual of the structured SVM, or max-margin Markov problem; partial decomposition via a gradient formulation; and finally tight coupling of a max-likelihood inference algorithm into the optimization algorithm, as opposed to using inference as a working set maintenance mechanism only. The tight coupling also allows fast approximate inference to be used effectively in the learning. The generality of the method follows from the fact that changing the output structure in essence only changes the inference algorithm, that is, the method can almost be used in 'plug and play' fashion.
This is joint work with Craig Saunders, Sandor Szedmak and John Shawe-Taylor.

Friday, July 14 (Room 106, Torrington Place, Roberts Building)

10:00–10:50: *Volker Tresp* (Siemens, Germany)
Learning informative nonparametric priors from multiple tasks

ABSTARCT TBA

11:00–11:50: *Ludovic Denoyer* (University of Paris 6, France)
XML structure mapping

A key problem for automating the processing of semi-structured resources is the format heterogeneity among data sources. For dealing with heterogeneous semi-structured data, the correspondence between the different formats has to be established. The multiplicity and the rapid growth of information sources have motivated researchers to develop machine learning technologies for helping to automate those transformations. In the Machine Learning community, structured output classification framework is an extension of the classical classification framework where outputs are distributed over an infinite discrete space. This general context encloses the tasks of sequence tagging, grammatical parsing, multiclass classification. Some interesting models have been developed in this framework (max margin networks, SVM ISO,.)
We focus here on the problem of mapping a document into a target mediated XML schema. We will first talk about the goals of this task. We will then explain how this task can be included in the Structured Classification framework and we will show why

existing models can't be used to solve this problem. We will finally propose a set of prospective models and explain the different results obtained.

14:00–14:50: *Marko Grobelnik* (J. Stefan Institute, Slovenia)
Going beyond bag-of-words: dealing with a text as a graph of triples

ABSTARCT TBA

15:00–15:50: *Theodoros Evgeniou* (INSEAD, France)
Learning multiple tasks with kernel methods

ABSTARCT TBA