Thomas Dietterich
*Divide-and-Conquer Methods for Spatio-Temporal Problems*

Many current and emerging machine learning applications involve learning a mapping from complex X values to complex Y values, where “both” the X’s and Y’s may be sequences, time series, images, GIS layers, or even sequences of images, sequences of GIS layers, etc. There are two conventional approaches to solving these problems: (a) methods based on graphical models such as HMMs and (b) divide-and-conquer methods. In the divide-and-conquer approach, each X and Y object is divided into a set of “windows” or “frames” such that window xi maps to output yi where yi can be treated as a class label over a small number of classes. Then standard classification methods can be applied to learn a window mapping yi = f(xi). Finally, some kind of “merging” operation combines the predicted yi values to obtain the entire object Y. This talk will briefly review the shortcomings of both the HMM and divide-and-conquer approaches and outline some research questions that I hope to answer as part of the ITR project.

Ashutosh Garg
*Learning Coherent Concepts*

Global, task specific constraints, can often make the learning problem easier as compared to what is suggested by existing learning theory. Co-existence of multiple learners can be modeled as a coherency constraint which has been shown to ease the problem of learning half-spaces. A General framework for modeling the constraints is presented. Various other kinds of constraints (class bias, data's existence in lower dimensional space) are being analyzed and their influence on the effective VC-dimension is being studied.

David Gondek
*A Latent-Variable Stochastic Model for the Web Graph*

Existing random graph generation models designed to exhibit the statistics found in the web graph have largely been defined over observable quantities, i.e. the in-degree/out-degree of a vertex, the number of vertices in the graph, etc. We propose a stochastic process for generating graphs which uses latent information. Specifically, we associate with each vertex a feature vector, assumed to be sparse. Motivated by psychometric models of additive clustering, edges are then assigned to vertices based on an additive function over their feature vectors' shared components. This generative process results in a graph exhibiting “overlapping clusters”. We discuss these graphs and metrics for determining whether a given graph exhibits such overlapping clusters.

Thomas Hofmann
*Information Retrieval as Recommendation*

Most information retrieval and filtering systems are based on a representation and (often rudimentary) analysis of document content. On the other hand, recommender systems are typically based on user profile data on which collaborative filtering is performed. In this talk, I will discuss ways to systematically combine content-based filtering with collaborative filtering for personalized information retrieval.
**Vasin Punyakanok**  
*Inference with Classifiers*

We study the problem of combining the outcomes of several different classifiers in a way that provides a coherent inference that satisfies some constraints. In particular, we develop two general approaches for an important subproblem - identifying phrase structure. The first is a Markovian approach that extends standard HMMs to allow the use of a rich observation structure and of general classifiers to model state-observation dependencies. The second is an extension of constraint satisfaction formalisms. We develop efficient combination algorithms under both models and study them experimentally in the context of shallow parsing.

**Jason Rennie & Chen-Hsiang Yeang**  
*An Information Theoretic Framework for Multi-class Classification Using Output Codes*

Building the code matrix for the error-correcting output coding (ECOC) scheme is one of the challenging problems in multi-class categorization. In this paper we present an information theoretic framework of building the code matrix for the ECOC scheme. It incorporates partitions incrementally by maximizing the mutual information between target labels and classification results. Using mutual information has a number of advantages including a method for balancing the desire for accurate classifiers with the need for classifiers to be relatively independent and distinguish between different aspects of the data. In this talk, we develop an iterative algorithm for creating a code matrix based on mutual information, discuss the advantages over other techniques and present some preliminary results.

**Rif Rifkin and Sayan Mukherjee**  
*Multiclass Classification with Support Vector Machines*

We are interested in solving multiclass classification problems by combining the outputs of binary classifiers. We use a Support Vector Machine as our underlying binary classifier. We explore one-vs.-all, all-pairs and error-correcting code schemes. On a fourteen class tumor morphology discrimination task, we find that the one-vs.-all scheme performs best, and indeed performs as well or better than any other classification scheme tried.

**Daniel Roth**  
*Relational Learning via Propositional Algorithms*

I will discuss a new paradigm for relational learning which supports representing and learning relational information using propositional means. This paradigm suggests different tradeoffs than those in the traditional approach to this problem -- the ILP approach -- and as a result it enjoys several significant advantages over it. Examples will be given from NLP and information extraction tasks.

**Thomas Serre**  
*Component-based Face Detection*

We present a component-based, trainable system for detecting frontal and near-frontal views of faces in still gray images. The system consists of a two-level hierarchy of Support Vector Machine (SVM) classifiers. On the first level, component classifiers independently detect components of a face. On the second level, a single classifier checks if the geometrical configuration of the detected components in the image matches a geometrical model of a face.
Sebastian Seung  
*Nonnegative Matrix Factorization for Feature Discovery*

I will discuss some theoretical and practical aspects of applying nonnegative matrix factorization to the problem of learning features.

Pawan Sinha  
*Recognition of Impoverished Images: Making do with very Little*

I will describe recent experimental studies that explore how well humans can recognize highly degraded images. An applied goal is to determine how little image information should, in principle, suffice for an artificial system to perform similar recognition tasks.

Martin Szummer  
*Kernel Expansions with Unlabeled Examples*

Modern classification applications necessitate supplementing the few available labeled examples with unlabeled examples to improve classification performance. We present a new tractable algorithm for exploiting unlabeled examples in discriminative classification. This is achieved essentially by expanding the input vectors into longer feature vectors via both labeled and unlabeled examples. The resulting classification method can be interpreted as a discriminative kernel density estimate and is readily trained via the EM algorithm, which in this case is both discriminative and achieves the optimal solution. We provide, in addition, a purely discriminative formulation of the estimation problem by appealing to the maximum entropy framework. We demonstrate that the proposed approach requires very few labeled examples for high classification accuracy.

Kinh Tieu  
*Boosting Image Retrieval*

We present an approach for image retrieval using a very large number of highly selective features and efficient online learning. Our approach is predicated on the assumption that each image is generated by a sparse set of visual “causes” and that images which are visually similar share causes. We propose a mechanism for computing a very large number of highly selective features, which capture some aspects of this causal structure (in our implementation there are over 45,000 highly selective features). At query time a user selects a few example images, and a technique known as “boosting” is used to learn a classification function in this feature space. By construction, the boosting procedure learns a simple classifier that only relies on 20 of the features. As a result a very large database of images can be scanned rapidly, perhaps a million images per second. Finally we will describe a set of experiments performed using our retrieval system on a database of 3000 images.

Florian Wolf  
*Existing Research on Document Classification by Humans*

Research on document classification by humans suggests that the strategies employed in these tasks are highly flexible. Which strategy is employed seems to be influenced in particular by expertise, task goal, document complexity, and document structure. Furthermore, existing research suggests that a basic heuristic of humans for document classification is to rely on those document features that are most informative, given a certain task goal and kind of document. This presentation will review these findings and make suggestions for further experimental research that examines in more detail how reliable “shallow features” (such as document layout) are for document classification as opposed to “deep features” (such as propositional content).