

NSF-ITR/IM PROJECT: 2003 Abstracts

From Bits to Information: Statistical Learning Technologies for Digital Information Management Search (Brown University, M.I.T, Oregon State University, and University of Illinois at Urbana-Champaign)

Brown University:

Project Title: Support Vector Machines for Multiple Instance Learning *PI: T. Hofmann*
Participants: Stuart Andrews and Thomas Hofmann
Abstract: Multiple Instance Learning (MIL) is an important generalization of standard supervised binary classification. In MIL labels are not available for individual training patterns, but are associated with sets of patterns, thereby introducing ambiguity in the correspondence between patterns and labels. This project investigates advanced optimization techniques such as disjunctive programming and reformulation linearization techniques to generalize large margin methods such as Support Vector Machines and Linear Programming Boosting to the MIL case. This is accomplished by convexifying the non-convex margin constraints that are introduced by the training data ambiguity. On the application side our main goal is to apply these techniques to multi-class object detection and recognition problems to enable practical, flexible and efficient content-based retrieval of images.

Project Title: Polycategorical Categorization for Personalized Information Filtering *PI: T. Hofmann*
Participants: Ioannis Tsochantaridis and Thomas Hofmann
Abstract: This project investigates how to extend multi-class classification methods to learn mappings to arbitrary discrete output spaces. The goal is to considerably extend the applicability of machine learning methods and to provide a common framework for problems such as, label sequence learning (i.e. annotating and segmenting observation sequences), classification with class taxonomies and more generally with classes characterized by features, and learning mappings to strings or trees. This is accomplished in the proposed framework by allowing features to be extracted jointly from inputs and outputs (so-called joint features) and by learning a linear discriminant in this joint feature space by standard large margin methods, thereby generalizing multi-class Support Vector Machines to this more general setting.

Massachusetts Institute of Technology:

Project title: Representation and Detection of Shapes in Images *PI: E. Grimson*
Participant: Pedro F. Felzenszwalb
Abstract: We are developing techniques to represent and detect two-dimensional shapes in images. Besides describing shapes in a natural and precise manner, our representation can characterize generic shapes at multiple levels of specificity. Moreover, the representation yields fast algorithms for various computational tasks. For example, we have developed an algorithm that can efficiently localize a non-rigid object in an image. A second algorithm can be used to detect potential objects in images without using a specific shape model. We believe these algorithms will be useful for content based image classification and retrieval.

Project Title: An Online Linear Prediction Model For Example-driven Image Retrieval *PI: E. Grimson*
Participants: Huizhen Yu and Eric Grimson
Abstract: In content-based image or multimedia retrieval, usually the items have several qualitatively different types of features, such as text annotations, color, visual texture, structured patterns etc. On each of those feature spaces the system has expert submodules that can generate a classifier based on examples given by the user. Given a few examples, it is challenging to quickly identify relevant feature spaces and track the user's possibly shifting interests. Most previous approaches treat retrieval as a batch-learning problem on a joint vector feature space and optimize the goodness-of-fit on the examples, thus neglecting the temporal order in the data and are prone to overfitting. We show an online linear prediction model with an associated class of online learning algorithms, which is more suitable for this example-driven retrieval problem. It adaptively weights the retrieval submodules based on their forward prediction errors, thus avoiding overfitting of the combined classifier. In cases where a user's interest shifts, variants of weighting schemes can be used to automatically track the best submodule over time. Our experiments show the great potential of applying the online learning framework to image retrieval. We view this as an intermediate step beyond previous batch learning work to model retrieval as a more complex online decision problem.

Project Title: **A Compression Framework for Text Classification** *PI: T. Jaakkola*

Participants: Jason Rennie and Tommi Jaakkola

Abstract: The problem of text classification is to predict the labels of documents based on their content. When learning a classifier using training data, one must balance two aspects: (1) specificity to the training set, and (2) generalization to unseen data. A natural framework for this problem is compression. In compressing a set of document labels, we can either transmit the labels directly, or specify aspects of the documents that predict the labels. This imposes a natural trade-off: it is advantageous to use features from the documents only if they allow more efficient encoding of the labels. Important in bridging the gap between theory and practice is efficiently encoding features and weights. We develop techniques for this by taking advantage of prior information. The compression framework also naturally extends to incorporating unlabeled data, making use of similar classification problems, and taking advantage of hierarchical structure. We describe these extensions and provide examples on real-world data sets.

Project Title: **Generalized Low-Rank Approximations** *PI: T. Jaakkola*

Participants: Nathan Srebro and Tommi Jaakkola

Abstract: We study the frequent problem of approximating a target matrix with a matrix of lower rank. We provide a simple and efficient (EM) algorithm for solving /weighted/ low-rank approximation problems, which, unlike simple matrix factorization problems, do not admit a closed form solution in general. We analyze, in addition, the nature of locally optimal solutions that arise in this context, demonstrate the utility of accommodating the weights in reconstructing the underlying low rank representation, and extend the formulation to non-Gaussian models such as logistic models. We apply the methods developed to a collaborative filtering task.

Project Title: **Information Regularization with Partially Labeled Data** *PI: T. Jaakkola*

Participants: Martin Szummer (CBCL, MIT AI Lab) and Tommi Jaakkola (MIT AI Lab)

Abstract: Classification with partially labeled data requires using a large number of unlabeled examples, or $P(\mathbf{x})$, to further constrain the conditional $P(y|\mathbf{x})$ beyond a few available labeled examples. We formulate here a regularization approach to linking the marginal and the conditional in a general way. The regularization penalty measures the information that is implied about the labels at the chosen resolution over $\{\epsilon_m \text{ covering}\}$ regions. No parametric assumptions are required and the approach remains tractable even for continuous marginal densities $P(\mathbf{x})$. We develop algorithms for solving the regularization problem for finite covers, establish a limiting differential equation, and exemplify the behavior of the new regularization approach in simple cases.

Project Title: **On Information Regularization** *PI: T. Jaakkola*

Participants: Adrian Corduneanu and Tommi Jaakkola

Abstract: We formulate a principle for classification with the knowledge of the marginal distribution over the data points (unlabeled data). The principle is cast in terms of Tikhonov style regularization where the regularization penalty articulates the way in which the marginal density should constrain otherwise unrestricted conditional distributions. Specifically, the regularization penalty penalizes any information introduced between the examples and labels beyond what is provided by the available labeled examples. The project extends the work of Szummer and Jaakkola (2002) to multiple dimensions, providing a regularizer independent of the covering of the space used in the derivation. We illustrate the regularization principle in practice by restricting the class of conditional distributions to be logistic regression models and constructing the regularization penalty from a finite set of unlabeled examples.

Project Title: **Component-based Face Detection** *PI: T. Poggio*

Participants: Bernd Heisele and Thomas Serre

Abstract: 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 experts 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. We propose a method for automatically learning components by using 3-D head models. This approach has the advantage that no manual interaction is required for choosing and extracting components. Experiments show that the component-based system is significantly more robust against rotations in depth than a comparable system trained on whole face patterns.

Project Title: **Classification of Yahoo News from Images and Captions** *PI: T. Poggio*

Participants: Alexandros Kyriakides and Giorgos Zacharia

Abstract: Classification of images obtained from news articles. Our goal is to find the best suitable image for a news story. We use the captions of the images in order to classify the images and perform the selection. Our current approach tries to predict which words in a news story are likely to be in the caption of the image. In essence we are selecting the most important words from a news story that will help us select the correct image. We can currently predict with 75%-80% accuracy if a word in a news story is also in the caption of the image. The confusion matrix

below shows how well one of our predictors performs. Words, which are both in the news story and the caption, are termed "yes" words. Words, which are in the news story but not in the caption, are termed "no" words. The rows indicate the actual class. The columns indicate the predicted class.

	No	Yes
no	120618	32040
yes	4771	8764

Correctly Classified Instances: 129382 (77.8505 %)

Project Title: Stability is necessary and sufficient for consistency of Empirical Risk Minimization

Participants: Sayan Mukherjee and Tomaso Poggio *PI: T. Poggio*

Abstract: Solutions of learning problems by Empirical Risk Minimization (ERM) -- and almost-ERM in general when the minimizer does not exist -- need to be *consistent*, so that they may be predictive. They also need to be *well-posed*, so that they might be used robustly. We propose a statistical form of well-posedness, defined in terms of the property of *cross-validation leave-one-out cv stability*. Our main observation implies that for bounded loss classes, *cv stability of ERM is necessary and sufficient for consistency of ERM*. We conclude that *cv stability* is the weakest form of stability which is sufficient for convergence in probability of the empirical error to the expected error for general learning algorithms while being necessary and sufficient for ERM. We discuss stronger forms of stability and their relations with "small" uGC hypothesis spaces, such as VC-classes and balls in a Sobolev or a RKHS space.

Project Title: Rademacher Averages as Complexity Measures

PI: T. Poggio

Participant: Alex Rakhlin

Abstract: We study measures of complexity other than VC or V-gamma dimensions. In many applications, the Rademacher Averages appear to be an easier tool than the usual metric entropy approaches. In particular, we show that using recent developments in Empirical Process theory, we get fast convergence rates for estimating densities from very rich classes of infinite convex combinations of parametrized densities.

Project title: Information Regularization with Partially Labeled Data

PI: T. Poggio

Participants: Martin Szummer and Tommi Jaakkola

Abstract: Classification with partially labeled data involves learning from a few labeled examples as well as a large number of unlabeled examples, and represents a blend of supervised and unsupervised learning. We formulate a regularization approach to linking the marginal density (estimated from unlabeled data) to the conditional (estimated from labeled data) in a general way. The regularization penalty measures the information that is implied about the labels over covering regions. No parametric assumptions are required and the approach remains tractable even for continuous marginal densities $P(x)$. We develop algorithms for solving the regularization problem for finite covers, establish a limiting differential equation, and exemplify the behavior of the new regularization approach.

Project Title: Generalized Robust Conjoint Estimation

PI: T. Poggio

Participants: Giorgos Zacharia, Constantinos Boussios, Theodoros Evgeniou

Abstract: We present a framework within which models of preferences that are robust to noise and can be highly nonlinear are computationally efficiently estimated. The models are in the spirit of recently developed polyhedral methods for conjoint analysis, and they bring together ideas from statistical learning theory as well as optimization theory to the field of preference modeling and conjoint analysis. We compare these models with standard logistic regression, Hierarchical Bayes, and the polyhedral conjoint estimation methods using standard, widely used, simulation data. The experiments show that the proposed methods can handle noise better than existing methods and can be used for fast and better estimation of nonlinear preference models. They can therefore be useful, for example, for analyzing data that are noisy such as that describing universal choices or clicks on the Internet, or for estimating interactions among product features.

Project Title: STICKS: Image-representation via non-local comparisons

PI: P. Sinha

Participants: Benjamin J. Balas and Pawan Sinha

Abstract: A fundamental question in visual neuroscience is how to represent image structure. The most commonly used representation scheme relies on spatially localized differential operators, approximated as Gabor filters with a set of excitatory and inhibitory lobes, which compare adjacent regions of an image. While well-suited to encoding local relationships, such operators have some significant drawbacks. Specifically, they confound a filter's inter-lobe distance with the size of the lobes themselves. Thus, to make comparisons across larger image distances, the scheme uses filters with larger lobes, which implies spatial averaging over larger areas. This leads to problems when one tries to compare small regions across large distances. In order to address this problem, we introduce the "dissociated dipole" or "sticks" operator, for performing non-local comparisons within an image. This operator decouples lobe size from inter-lobe distance and allows for parametric movement between edge-based and region-based representation modes. Here we report on two aspects of sticks. First, we assess perceptual plausibility of the

operator via psychophysical experiments that test observers' ability to compare brightness of small target regions across large distances in an image. Our results suggest that subjects' thresholds are remarkably robust even over large separations (~15 degrees of visual angle) of the target regions. Second, to evaluate the effectiveness of this approach for image encoding, we have implemented a sticks-based system for content-based image retrieval. We have obtained good results across a diverse set of domains including outdoor scenes, faces and letters. Furthermore, performance appears to be robust against significant degradations such as resolution loss and occlusions. Based on these results, we believe that the sticks operator can serve as an effective scheme for representing image structure. Furthermore, this representation strategy may be useful for content retrieval tasks in non-visual domains as well, such as speech.

Project Title: Human Document Classification Using Bags of Words

PI: P. Sinha

Participants: Florian Wolf, Tomaso Poggio and Pawan Sinha

Abstract: Humans are remarkably adept at classifying text documents into categories. For instance, while reading a news story, we are rapidly able to assess whether it belongs to the domain of finance, politics or sports. Automating this task can have many practical applications such as content-based search or filtering of digital documents. To this end, it is interesting to investigate the nature of information humans use in order to accomplish document classification tasks. Here we examine whether classification can be performed even in the absence of syntactic and layout information. We have developed a novel paradigm of progressive revealing, which allows us to determine classification performance as a function of number of words seen with or without the maintenance of appropriate syntax. We find that although performance increases more rapidly with number of words when syntactic information is available, subjects are able to achieve similar classification accuracy even without syntactic information beyond a modest passage size. These results have implications for models of human text-understanding. They also allow us to estimate for a given passage size what level of performance we can expect, in principle, from a system without requiring a prior step of complex natural language processing.

Project Title: A Computational Scheme for Filling-in Missing Information in Images and Its Neural Correlates

PI: P. Sinha

Participants: Ethan Meyers, Yuri Ostrovsky and Pawan Sinha

Abstract: Much of visual processing can be characterized as the filling-in of missing information. For example, noise reduction involves the filling in of missing picture information. 3D shape recovery involves filling in the missing third dimension, and object recognition often requires implicit filling in of information behind occluders. Here we develop a simple computational scheme that can predict missing information in a variety of settings. Our method relies on learning non-local statistical dependencies between different image regions from training data. The notion of 'regions' we use is broad and includes not only spatially distinct sub-images, but also various sections of spatial frequency space. The inter-region dependencies are used to predict missing values in novel images. Because of the broad definition of regions, this approach is able to perform seemingly very different filling-in tasks including compensating for scotomas (filling-in in the spatial domain) and resolution enhancement (filling-in in the frequency domain). More generally, this scheme is a way to computationally formalize intuitions about the knowledge-based and non-local context-based strategies the brain uses to compensate for image degradations. In accompanying work, we have examined possible neural correlates of filling-in. Specifically, using the techniques of fMRI and magneto-encephalography, we have tested whether object specific neural activity can be elicited even when the intrinsic information about an object is almost completely missing and has to be inferred (filled-in) by contextual cues. We find that activity in the fusiform gyrus and at the occipito-temporal border exhibits this characteristic. This is the first demonstration of contextually elicited neural activity and serves as a starting point for elucidating the neural correlates of the various stages of our computational model of filling-in.

Oregon State University:

Project Title: Supervised Reinforcement Learning

PI: T. Dietterich

Participants: Xin Wang, Tom Dietterich

Abstract: In earlier work, Wei Zhang showed that reinforcement learning could be applied to learn application-specific evaluation functions for combinatorial optimization problems in industry, specifically, resource-constrained scheduling in NASA's space shuttle program. We are exploring ways to generalize his results to the problem setting that we call Supervised Reinforcement Learning. In this setting, a series of training examples is provided for a sequential decision-making task (such as scheduling). The goal is to learn an evaluation function that performs well on these examples and that also generalizes well to new instances. We have developed new kernel-based reinforcement learning algorithms for this setting and applied them to problems of scheduling and deterministic control. However, these methods are not fully satisfactory, so we are exploring a new model-based policy gradient method. In this method, a model of the reinforcement-learning problem is acquired through experimentation, the

behavior of a parameterized policy on this model is then employed to compute the gradient of the policy performance with respect to its parameters. We have obtained excellent results on several benchmark problems.

Project Title: **Content-Based Image Retrieval of Herbarium Samples** *PI: T. Dietterich*

Participants: T. Dietterich, Ashit Gandhi, Pengcheng Wu, Shriprakash Sinha

Abstract: Imagine you are hiking in the forest and you see an unusual plant. You want to know the genus and species of this plant, so you clip off a leaf, take it home, and scan it into your computer. Now you want to retrieve images of similar plants from online plant databases such as those being constructed by the Missouri Botanical Garden, the NY Botanical Garden, and the Oregon State Herbarium. We are developing an initial test bed for this problem based on 6 species of maples and oaks native to Oregon. We have developed a shape-based matching algorithm that borrows dynamic programming methods from DNA sequence matching and applies them to matching the sequence of local curvatures of the shape of a leaf. This allows us to match partial, overlapping, and occluded leaves successfully. This gives good results, but we are trying to improve them by using the dynamic programming distance measure as a kernel in a support vector machine. An interesting problem arises here, because the training data (Herbarium samples) is different from test data (isolated leaves). We are exploring two approaches to handle this problem, both of which exploit the availability of a small number of isolated leaves during training. First, we match the isolated leaves against the Herbarium samples to find chunks of the Herbarium samples that best resemble isolated leaves. We then extract these chunks and provide them as input to training examples to the SVM. Second, we have formulated the SVM to train on both isolated and Herbarium samples, but to constrain the support vectors to consist only of Herbarium samples.

Project Title: **Automatic Extraction of Label Data from Herbarium Samples** *PI: T. Dietterich*

Participants: T. Dietterich, Brian Breck (undergraduate)

Abstract: Each herbarium sample consists of a large dried, pressed sample of a plant, which is attached to a large sheet of cardboard. On this sheet there is also a label (handwritten, typewritten, or printed), that provides information such as the genus, species, name of collector, place found (including county), date collected, associated species, elevation, and so on. As discussed in the ITR proposal, we seek to construct information retrieval systems that can combine image and textual information. To do this, we want to capture this textual information and populate a database for the Herbarium. Approximately 50,000 samples labels have been manually transcribed into the database, but an additional 100,000 remain to be done. We are applying a commercial OCR program to read the typewritten and printed labels. Then we will develop spelling correction methods for fixing OCR errors. Finally, we plan to apply sequential supervised learning methods (see below) to extract the database fields from the spelling-corrected text.

Project Title: **Machine Learning for Sequential and Spatial Data** *PI: T. Dietterich*

Participants: T. Dietterich, Adam Ashenfelder, Saket Joshi

Abstract: The goal of this project is to build off-the-shelf algorithms for classifying sequences and spatial objects (e.g., pixels in images). Existing approaches, such as hidden Markov models and Markov Random Fields, require substantial hand-tweaking to work well in new applications. We seek discriminative methods that work well across a wide range of applications without manual tuning. We are studying two approaches to building such general off-the-shelf methods. The first approach is based on the Conditional Random Field defined by Lafferty, *et al.* We have developed a gradient boosting algorithm for fitting the parameters of the CRF as a weighted sum of regression trees. The resulting algorithm gives excellent results and is orders of magnitude faster than applying standard conjugate-gradient methods. The second approach is to learn recurrent classifiers. We have developed a general recurrent classifier implementation and integrated it into the WEKA machine-learning environment. We are currently conducting a study comparing several base-level learning algorithms within this recurrent framework on several interesting sequential data problems.

Project Title: **Sub-pixel Classification of Remote Sensed Images** *PI: T. Dietterich*

Participants: T. Dietterich, Diane Damon

Abstract: Many satellite-based instruments collect earth-surface images at coarse resolution (e.g., one pixel = 1km square or 0.5km square). Training data is available at higher resolutions, and sub-pixel classification attempts to predict the fraction of each pixel that belongs to various land-cover classes (forest, grassland, bare soil, crops, city, etc.). We have developed a new regression tree method for sub-pixel classification. Experiments show that it is more accurate (and easier to implement and interpret) than current state-of-the-art methods such as linear unmixing and neural networks.

Project Title: **Bias-Variance Analysis of SVM Classifiers** *PI: T. Dietterich*

Participants: G. Valentini (Univ. of Genoa), T. Dietterich

Abstract: A fundamental question is whether support vector machines can benefit from this (or any other) form of ensemble learning. We have conducted a bias/variance analysis of SVMs which shows that SVMs trained on

small samples can exhibit substantial variance, which suggests that variance-reduction methods, such as bagging, will be able to improve SVM performance. We have developed a strategy, called LoBag, which tunes SVMs to have low bias and then applies bagging to reduce variance. LoBagging gives better results than well-tuned single SVMs and bags of such SVMs.

University of Illinois at Urbana-Champaign:

Project Title: **Constraint Classification for Multiclass Classification and Ranking** *PI: D. Roth*

Participants: Sarel Har Peled and Dav Zimak

Abstract: The constraint classification framework captures many flavors of multiclass classification including winner-take-all multiclass classification, multi-label classification and ranking. We present a meta-algorithm for learning in this framework that learns via a single linear classifier in high dimension. We discuss distribution independent as well as margin-based generalization bounds and present empirical and theoretical evidence showing that constraint classification benefits over existing methods of multiclass classification.

Project Title: **Feature Description Logic** *PI: D. Roth*

Participant: Chad Cumby

Abstract: We present a paradigm for efficient learning and inference with relational data using propositional means. The paradigm utilizes description logics and concepts graphs in the service of learning relational models using efficient propositional learning algorithms. We introduce a Feature Description Logic (FDL) - a relational (frame based) language that supports efficient inference, along with a generation function that uses inference with descriptions in the FDL to produce features suitable for use by learning algorithms. These are used within a learning framework that is shown to learn efficiently and accurately relational representations in terms of the FDL descriptions. The paradigm was designed to support learning in domains that are relational but where the amount of data and size of representation learned are very large; we exemplify it here, for clarity, on the classical ILP task of learning family relations. This paradigm provides a natural solution to the problem of learning and representing relational data; it extends and unifies several lines of works in KRR and Machine Learning in ways that provide hope for a coherent usage of learning and reasoning methods in large-scale intelligent inference.

Project Title: **On Kernel Methods for Relational Learning** *PI: D. Roth*

Participant: Chad Cumby

Abstract: Kernel methods have gained a great deal of popularity in the machine learning community as a method to learn directly in high-dimensional feature spaces. Those interested in relational learning have recently begun to cast learning from structured and relational data in terms of kernel operations. We attempt to study the benefits and drawbacks of kernel learning in relational domains by describing a general family of kernel functions built up from a description language of limited expressivity. This allows us to examine issues of time complexity in terms of learning with these and other relational kernels, and how these issues relate to generalization ability. Learning with kernels in this family directly models learning over an expanded feature space constructed using the same description language. We highlight the tradeoffs between using kernels in a very high dimensional implicit space versus a restricted feature space, through experiments in the domains of bioinformatics and natural language processing.

Project Title: **Learning a Sparse, Part-Based Representation for Object Detection** *PI: D. Roth*

Participant: Shivani Agarwal

Abstract: This project develops an approach for learning to detect objects in still gray images via a sparse, part-based representation. A vocabulary of information-rich object parts is automatically constructed from a set of sample images of the object class of interest. Images are then represented using parts from this vocabulary, together with spatial relations observed among them. Based on this representation, a feature-efficient learning algorithm learns to accurately detect instances of the object class. The method can be applied to any object with distinguishable parts in a relatively fixed spatial configuration. We have evaluated it on a difficult set of real-world images containing side views of cars, and find that it successfully detect objects in varying conditions amidst background clutter and partial occlusion. In evaluating object detection approaches, several important methodological issues arise that have not been satisfactorily addressed in previous work; in our work we also propose solutions to a number of these issues.

Project Title: **Automatic Learning of Feature Transformations for Pattern Classification** *PI: D. Roth*

Participant: Shivani Agarwal

Abstract: The question of how to learn good features for classification is an important one in machine learning. In this work, we formulate a framework for automatically learning the optimal feature transformation for a given classification problem, with respect to a given classification algorithm. The framework is based on extending the principle of risk minimization, commonly used for learning classifiers, to learning feature transformations that admit

classifiers with minimum risk. This allows feature extraction and classification to proceed in an integrated manner. The framework is applied to derive new algorithms for learning feature transformations; preliminary experiments demonstrate the ability of the resulting algorithms to learn good features for a variety of classification problems.

Project Title: On Generalization Bounds, Projection Profile, and Margin Distribution

PI: D. Roth

Participant: Ashutosh Garg

Abstract: We study generalization properties of linear learning algorithms and develop a data dependent approach that is used to derive generalization bounds that depend on the margin distribution. Our method uses random projection techniques to allow the use of existing VC dimension bounds in the effective, lower, dimension of the data. Our bounds are tighter than existing bounds and (sometimes) give informative generalization bounds for real world, high dimensional problems. We use these results to study new a practical data-dependent complexity measure for learning. The new complexity measure is a function of the observed margin distribution of the data, and can be used, as we show, as a model selection criterion. This is used to develop the Margin Distribution Optimization (MDO) learning algorithm that directly optimizes this complexity measure. Empirical evaluation of MDO demonstrates that it consistently outperforms SVM.

Project Title: Learning Coherent Concepts

PI: D. Roth

Participants: Ashutosh Garg and Vasin Punyakanok

Abstract: This research seeks to develop an integrated view - theoretical understanding, algorithms development and experimental evaluation - for learning coherent concepts. These are learning scenarios that are common in cognitive learning - where multiple learners co-exist and may learn different functions on the same input, but there are mutual compatibility constraints on their outcomes. Our effort will consist of developing a learning theory for these situations and of studying algorithmic ways to exploit them in natural language inferences. The theoretical study concentrates on developing a semantics for the coherency conditions and study it from a learning theory point of view. The goal is to understand in what ways does learning become easier and more robust in these situations. The algorithmic study concentrates on developing ways to exploit coherency and makes use of several important problems in natural language processing as a testbed for investigating chaining of coherent classifiers and inferences that rely on the outcomes of several classifiers.

Project Title: Constraint Classification: A New Approach to Multiclass Classification

PI: D. Roth

Participants: Dav Zimak and Yair Even-Zohar

Abstract: We develop a new view of multiclass classification and introduce the constraint classification problem, a generalization that captures many flavors of multiclass classification. In particular, our framework captures multiclass classification, ranking problems and multi-label classification and winner-take-all (WTA) algorithms. We study both algorithmic issues and theoretical issues such as sample bounds. Algorithmically, based on our view, we develop a learning algorithm that learns via a single linear classifier in high dimension and can also be viewed as a network of properly trained linear classifiers in a low dimension. We also study distribution independent bounds for many multiclass-learning algorithms, including winner-take-all (WTA), as well as margin-based generalization bounds.

Project Title: Learning Sparse Representations for Object Detection

PI: D. Roth

Participants: Shivani Agarwal and Ashutosh Garg

Abstract: We study an approach for learning to detect objects in still gray images that is based on a sparse, part-based representation of objects. A vocabulary of information-rich object parts is automatically constructed from a set of sample images of the object class of interest. Images are then represented using parts from this vocabulary, along with spatial relations observed among them. Based on this representation, a feature-efficient learning algorithm is used to learn to detect instances of the object class. The framework developed can be applied to any object with distinguishable parts in a relatively fixed spatial configuration. So far we have experimented on images of side views of cars. Our experiments show that the method achieves a high-detection accuracy on a difficult test set of real-world images, and is highly robust to partial occlusion and background variation.

Project Title: Intermediate Knowledge Representations that Facilitate Learning

PI: D. Roth

Participants: Dav Zimak, Chad Cumby and Shivani Agarwal

Abstract: Learning becomes easy once the correct input representation has been chosen, for example, one that produces linearly separable point sets. We have several projects in the direction of: (1) automatically generating intermediate representations to aid supervised learning algorithms, (2) developing methods that allow the use of relational representations and of learning relational definitions, and (3) developing a flexible knowledge representation language that can be used along with feature efficient learning algorithms. We study applications of this general knowledge representation paradigm in the context of learning in the natural language domain (e.g., information extraction) and visual recognition. (4) Developing kernels for Boolean functions and relational functions and studying computational complexity of algorithms that use kernels.

Project Title: **Inference with Classifiers**

PI: D. Roth

Participants: Dan Roth and Vasin Punyakanok

Abstract: In many situations it is necessary to make decisions that depend on the outcomes of several different classifiers in a way that provides a coherent inference that satisfies some constraint. These constraints might arise from the sequential nature of the data or other domain specific constraints. We study several general approaches to this problem and are evaluating those in the context of inference problems in natural language -- identifying phrase structure and question/answering. The approaches studied are: (1) 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. We study both generative and conditional models. (2) Extensions of constraint satisfaction formalisms. Currently the focus is on developing hierarchical models. (3) Markov Random fields. We study a more general model in which constraints of more general structures can be developed.