07161 Abstracts Collection
Probabilistic, Logical and Relational Learning - A Further Synthesis
— Dagstuhl Seminar —

Luc De Raedt[1], Thomas Dietterich[2], Lise Getoor[3], Kristian Kersting*[4] and Stephen H. Muggleton[5]

[1] Katholieke Universiteit Leuven, B
luc.deraedt@cs.kuleuven.be
[2] Oregon State Univ., US
tgd@cs.orst.edu
[3] Univ. of Maryland - College Park, US
getoor@cs.umd.edu
[4] MIT, CSAIL, US
kersting@csail.mit.edu
[5] Imperial College London, UK
shm@doc.ic.ac.uk

Abstract. From April 14 – 20, 2007, the Dagstuhl Seminar 07161 "Probabilistic, Logical and Relational Learning - A Further Synthesis" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Artificial Intelligence, Uncertainty in AI, Probabilistic Reasoning, Knowledge Representation, Logic Programming, Relational Learning, Inductive Logic Programming, Graphical Models, Statistical Relational Learning, First-Order Logical and Relational Probabilistic Languages

07161 Summary of 'Probabilistic, Logical and Relational Learning - A Further Synthesis'

Data Mining and Machine Learning are in the midst of a "structured revolution". After many decades of focusing on independent and identically-distributed (iid) examples, many researchers are now studying problems in which examples consist of collections of inter-related entities or are linked together. A major driving

* Seminar-Coordinator

force is the explosive growth in the amount of heterogeneous data that is being collected in the business and scientific world. Example domains include bioinformatics, chemoinformatics, transportation systems, communication networks, social network analysis, link analysis, robotics, among others. The structures encountered can be as simple as sequences and trees (such as those arising in protein secondary structure prediction and natural language parsing) or as complex as citation graphs, the World Wide Web, and logical knowledge bases. In all these cases, structured representations can give a more informative view of the problem at hand, which is often crucial for the development of successful mining and learning algorithms.

The field of *probabilistic, logical and relational learning* (aka. *statistical relational learning*, *probabilistic inductive logic programming*) tackles the structured input-output problem sketched above by combining expressive knowledge representation formalisms such as relational and first-order logic with principled probabilistic and statistical approaches to inference and learning, and hence lies at the heart of artificial intelligence. It is a relatively young and all the more active field of research offering a lot of research opportunities. This was already witnessed by a previous seminar on "Probabilistic, Logical and Relational Learning - Towards a Synthesis" that took place from January 30 to February 04, 2005, which succeeded in bringing together a significant number of researchers from all over the world that are working on all aspects of probabilistic, logical and relational learning. The result was a better understanding of the common grounds of this newly emerging field and the identification of a number of key research challenges.

The goal of the 2007 seminar was to provide answers to some of this key research challenges in the area, including:

1. What is the relationship among the many different probabilistic, logical and relational representations that are being used?
2. What are suitable settings for learning such representations? And, what are the challenges raised by the different learning settings? Also, can one arrive at a learning theory focused on probabilistic, logical and relational representations?
3. What are the application areas for which probabilistic, logical and relational learning is well-suited? What does it take to develop show-case applications in these areas? Can we identify common and concrete application challenges on which progress can be measured and techniques? Providing answers to these questions should — ultimately — provide the field with a commonly agreed upon framework as well as provide an application focus, which together could form the basis for further developments in the area.

Not all of the questions could have been answered yet but significantly progress has been made as shown by the great collection of abstract below. They have been collected from 45 seminar attendees from 11 different countries. The presentations at the seminar, varying in length, covered a large variety of topics, including novel results on lifted inference within first-order probabilistic languages,

learning infinite relational models, statistical predicate invention, and applications within citation analysis, robotics, and life sciences. Talks were spread over the week to allow for plenty of time for discussions. Breakout sessions on special interest topics were organized on the fly using the Seminar's Wiki page. The breakout sessions gave the participants a chance to exchange problems and discuss ideas and challenges lying ahead indepth. We are positive that many of the breakout sessions will lead to new results, collaborations, and publications. Within the talks and the breakout sessions, we saw very lively debates showing the growing demand and opportunities for statistical relational learning within theory and practice of machine learning. We were also very pleased to see the significant progress made between the present seminar and the previous one. This was very clear in the demonstration session, where a number of academic prototypes of probabilistic, logical and relational learning systems were presented.

As usual, Schloss Dagstuhl proved to be an excellent place to hold a great meeting, so we would not only like to thank the participants of the seminar for making this a very successful event, but also the Dagstuhl staff for providing a friendly and stimulating working environment. Finally, we would like to thank Sriraam Natarajan for his valuable help in collecting the abstracts and full-text contributions.

Joint work of: De Raedt, Luc; Dietterich, Thomas; Getoor, Lise; Kersting, Kristian; Muggleton, Stephen

See also:
Wiki

Training and Prediction in Probabilistic First-Order Logic

Aron Culotta (Univ. of Massachusetts - Amherst, USA)

I will discuss enhanced training and prediction algorithms for probabilistic first-order logic.

Existing training algorithms have difficulty coping with the combinatorial explosion of predicates that results from grounding first-order formulae. I will present an error-driven, ranking-based algorithm that enables efficient parameter estimation by instantiating predicates incrementally as needed.

Existing prediction algorithms rely on general-purpose solutions, such as stochastic search. However, many problems can be decomposed into a combination of subproblems, each of which may have an efficient exact prediction algorithm (e.g., a dynamic program). Using a single general-purpose algorithm ignores this structure. Instead, I propose a sparse message-passing algorithm that combines the output of several local prediction algorithms to construct an approximate global solution.

Exploiting decomposability when estimating probability distributions over possible worlds

James Cussens (University of York, GB)

A PRISM program defines a distribution over possible worlds and the PRISM system allows maximum likelihood estimation (MLE) of the parameters of such distributions from data using the EM algorithm. In this talk I will show how to exploit structure in the underlying logic program to speed up the MLE process. The starting point for this analysis is that any hierarchical model is easily represented as a PRISM program: a hierarchical model is defined by a hypergraph over a set of variables and the structure of this hypergraph is representable by a single definite clause. ('Hierarchical models' in the sense meant here are a special case of log-linear models, but a generalisation of Markov networks.) From this encoding it becomes obvious that any hierarchical model can be fitted using the EM algorithm as an alternative to the more familiar Iterative Proportional Fitting (IPF) algorithm. Missing data and 'structural zeroes' are easily handled using this encoding.

If the hypergraph can be 'decomposed' into (i.e. is the direct join of) two smaller hypergraphs, then the MLE problem can be similarly broken down into two separate smaller MLE problems. If the hypergraph is (fully) decomposable then MLE is possible by simple counting without recourse to any iterative procedures like IPF.

It follows that 'decomposable' PRISM programs permit MLE without recourse to EM. One way of showing this is to see that decomposable programs can be restructured to remove 'failures' in the generative process implicit in a PRISM program (this is connected to the fact that any decomposable model is equivalent to a Bayesian network). The key research problem is to develop a notion of 'first-order decomposition' where the advantages of decomposability can be exploited for models which generalise (propositional) hierarchical models.

Keywords: PRISM, possible worlds, hierarchical models, decomposability

Model equivalence of PRISM programs

James Cussens (University of York, GB)

The problem of deciding the probability model equivalence of two PRISM programs is addressed. In the finite case this problem can be solved (albeit slowly) using techniques from *algebraic statistics*, specifically the computation of elimination ideals and Gröbner bases. A very brief introduction to algebraic statistics is given. Consideration is given to cases where shortcuts to proving/disproving model equivalence are available.

Keywords: PRISM programs, model equivalence, model inclusion, algebraic statistics, algebraic geometry, ideals, varieties, Gröbner bases, polynomials

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1380

SRL for Personalized Medicine: Will I Have a Heart Attack from this Drug?

Jesse Davis (University of Wisconsin - Madison, USA)

We wish to present a significant application domain for SRL community. With the advent of electronic medical records, researchers could have ready access to clinical records, and possibly genetic data, for patients participating in various research studies. Access to this combination of data will allow researchers to pose and investigate the following types of questions. Can one develop a model to predicate the efficacy of a potential drug for a given individual? Can clinical and genetic data provide insight into which individuals will have adverse reactions to a drug?

The interesting challenges posed by this data include the amount of available information, its multi-relational nature and its design to optimize ease of data access and billing rather than learning and modeling. The contributions of this talk are to (1) describe the application domain and its potential, (2) explain the necessity of explicitly modeling uncertainty, and (3) present a case study with a real-world database. The specific case study application is to learn a statistical model to indicate which patients on Cox2 inhibitors, such as Vioxx and Celebrex, are at substantial risk for heart attack.

Keywords: Personalized Medicine, Pharmacogenetics, Clinical Data

Joint work of: Davis, Jesse; Page, David; Lantze, Eric, Strufy, Jan; Pessig, Peggy; Vidaillet, Humberto; Caldwell, Michael

ProbLog : A probabilistic prolog and its applications to link discovery in biological network mining

Luc De Raedt (Katholieke Universiteit Leuven, B)

ProbLog is a recently introduced probabilistic extension of Prolog [De Raedt, Kimmig, Toivonen, IJCAI 07]. A ProbLog program defines a distribution over logic programs by specifying for each clause the probability that it belongs to a randomly sampled program, and these probabilities are mutually independent. The semantics of ProbLog is then defined by the success probability of a query in a randomly sampled program. It has been applied to link mining and discovery in a large biological network. In the talk, I shall also discuss the theory compression task for ProbLog, which consists in finding a small theory that best explains a given set of examples, and indicate some future directions.

Joint work of: Kimmig, Angeligka; Toivonen, Hannu; Revoredo, Kate; Kersting, Kristian

Keywords: Statistical Relational Learning, Probabilistic Logic Learning

Full Paper:
http://www.cs.helsinki.fi/u/htoivone/pubs/IJCAI07-397.pdf

Experience with Markov Logic Networks in a Large AI System

Thomas Dietterich (Oregon State University, USA)

CALO is an integrated AI system that seeks to provide support for the modern knowledge worker. Led by SRI, the CALO project includes contributions from more than 25 research groups in the U.S. To integrate the various learning components, and to combine hand-written probabilistic rules with factual and learned knowledge, we implemented and deployed a Markov Logic system that we call the Probabilistic Consistency Engine (PCE). This talk will describe our experiences with the PCE in CALO. The PCE provides several functions. First, it integrates probabilistic predictions from various learned and hand-authored components to maintain a relational model of the user's work environment (projects, appointments, action items, files, folders, email messages, email folders, web pages, etc.). Second, it provides a general mechanism for probabilistic credit assignment, so that corrective feedback from the user can be routed back to the classifier responsible for an incorrect prediction. Third, it provides a mechanism for various forms of co-training and semi-supervised learning. The talk will also discuss future directions for the PCE.

Keywords: Markov Logic, Integrating Learning and Reasoning

Joint work of: Dietterich, Thomas; Uribe, Thomas; Bao, Xinlong; Brigham, Chris

Markov Logic in Infinite Domains

Pedro Domingos (University of Washington, USA)

Markov logic combines logic and probability by attaching weights to first-order formulas, and viewing them as templates for features of Markov networks. Unfortunately, in its original formulation it does not have the full power of first-order logic, because it applies only to finite domains.

Recently, we have extended Markov logic to infinite domains, by casting it in the framework of Gibbs measures. In this talk I will summarize our main results to date, including sufficient conditions for the existence and uniqueness

of a Gibbs measure consistent with an infinite MLN, and properties of the set of consistent measures in the non-unique case.

(Many important phenomena, like phase transitions, are modeled by non-unique MLNs.) Under the conditions for existence, we have extended to infinite domains the result in Richardson and Domingos (2006) that first-order logic is the limiting case of Markov logic when all weights tend to infinity. I will also discuss some fundamental limitations of Herbrand interpretations (and representations based on them) for probabilistic modeling of infinite domains, and how to get around them.

Finally, I will discuss some of the surprising insights for learning and inference in large finite domains that result from considering the infinite limit.

Keywords: Markov logic networks, Gibbs measures, first-order logic, infinite probabilistic models, Markov networks

Joint work of: Domingos, Pedro; Singla, Parag

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1381

On classification, ranking, and probability estimation

Peter Flach (University of Bristol, GB)

Given a binary classification task, a ranker is an algorithm that can sort a set of instances from highest to lowest expectation that the instance is positive. In contrast to a classifier, a ranker does not output class predictions Ű although it can be turned into a classifier with help of an additional procedure to split the ranked list into two. A straightforward way to compute rankings is to train a scoring classifier to assign numerical scores to instances, for example the predicted odds that an instance is positive. However, rankings can be computed without scores, as we demonstrate in this paper. We propose a lexicographic ranker, LexRank , whose rankings are derived not from scores, but from a simple ranking of attribute values obtained from the training data. Although various metrics can be used, we show that by using the odds ratio to rank the attribute values we obtain a ranker that is conceptually close to the naive Bayes classifier, in the sense that for every instance of LexRank there exists an instance of naive Bayes that achieves the same ranking. However, the reverse is not true, which means that LexRank is more biased than naive Bayes. We systematically develop the relationships and differences between classification, ranking, and probability estimation, which leads to a novel connection between the Brier score and ROC curves. Combining LexRank with isotonic regression, which derives probability estimates from the ROC convex hull, results in the lexicographic probability estimator LexProb.

Keywords: Ranking, probability estimation, ROC analysis, calibration

Joint work of: Flach, Peter; Matsubara, Edson

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1382

Query-time Entity Resolution

Lise Getoor (University of Maryland - College Park, USA)

Entity resolution is a critical component of data integration where the goal is to reconcile database references corresponding to the same real-world entities. Given the abundance of publicly available databases that have unresolved entities, we motivate the problem of quick and accurate resolution for answering queries over such 'unclean' databases. Since collective entity resolution approaches — where related references are resolved jointly — have been shown to be more accurate than independent attribute-based resolution, we focus on adapting collective resolution for answering queries. We propose a two-stage collective resolution strategy for processing queries. We then show how it can be performed on-the-fly by adaptively extracting and resolving those database references that are the most helpful for resolving the query. We validate our approach on two large real-world publication databases where we show the usefulness of collective resolution and at the same time demonstrate the need for adaptive strategies for query processing. We then show how the same queries can be answered in real time using our adaptive approach while preserving the gains of collective resolution.

Diffusion Learning Machines

Marco Gori (University of Siena, I)

Most of nowadays learning schemes neglect data structure and useful relationships amongst the training examples. On the opposite, there are plenty of learning tasks from, amongst others, chemistry, biology, pattern recognition, and data mining that can take advantage from structural representations and from a systematic exploitation of data relationships. In this talk, I propose to represent the data by graphs where the nodes contains real vectors so as to provide a structural representation of each pattern, and to adopt the same formalism for expressing data relationships.

I introduce a new model, called diffusion learning machine (DLM), whose decision mechanism relies on a diffusion information process through the whole graphical domain more than on the classic target prediction that associates values to single examples.

I give conditions under which the proposed model turn out to be a globally stable dynamical system which returns a unique value on the nodes of the graphical domain. Moreover, I prove that the function on the nodes that is returned is

continuously differentiable with respect to the learning parameters, which makes it possible to use the tradition optimization framework for learning and derive an efficient neural-like gradient computational scheme.

Depending on the choice of the mechanism to implement the diffusion process, DLMs turn out to generalize recursive neural networks as well as classic diffusion models described by the Laplacian of the graph. Interestingly, apart from cases of special symmetries, the strong dynamical restrictions to yield stability do not limit the computational power of DLMs, since I prove that DLM exhibits a universal approximation property in graphical domains.

I report very promising experimental results for problems taken from graph matching, bioinformatics, and Web link analysis.

Keywords: Diffusion learning machines, diffusion kernels, learning in graphical domains, link analysis, relational learning

Towards Transductive Ranking

Thomas Gärtner (Fraunhofer IAIS - St. Augustin, D)

Whenever labelled and unlabelled data come from different distributions ("instance shift"), supervised learning algorithms can arrive at arbitrarily bad hypotheses. Naturally, to cope with such situations, one has to take unlabelled data into account. In this work we investigate different strategies to develop transductive ranking algorithms robust with respect to instance shifts. We will present theoretical motivations as well as algorithmic approaches.

Keywords: Tansduction, Ranking, Instance Shift

Clustering and visualization of relational data

Barbara Hammer (TU Clausthal, D)

Neural clustering and visualization algorithms such as the self-organizing map (SOM) and neural gas (NG) constitute very intuitive methods with widespread applications ranging from web mining to bioinformatics. Recently, variants which extend the classical algorithms which work on vectorial data only towards more general data structures have been proposed.

This includes recurrent and recursive versions for time series and hierarchical tree structures as well as extensions of so-called batch clustering to more general similarities or dissimilarities.

The talk will give an overview about extensions of SOM and NG to general data structures including several applications and theoretical findings and it will give a short glimpse towards a new proposal by means of the relational dual of the above methods.

Keywords: Relational clustering, median clustering, recursive SOM models, kernel SOM

CA general framework for unsupervised preocessing of structured data

Barbara Hammer (TU Clausthal, D)

We propose a general framework for unsupervised recurrent and recursive networks. This proposal covers various popular approaches like standard self organizing maps (SOM), temporal Kohonen maps, resursive SOM, and SOM for structured data. We define Hebbian learning within this general framework. We show how approaches based on an energy function, like neural gas, can be transferred to this abstract framework so that proposals for new learning algorithms emerge.

Keywords: Relational clustering, median clustering, recursive SOM models, kernel SOM

Joint work of: Hammer, Barbara; Micheli, Alessio; Sperduti, Alessandro

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1383

Type Extension Trees for Feature Construction and Learning in Relational Domains

Manfred Jaeger (Aalborg University, DK)

We introduce type extension trees as a formal representation language for complex combinatorial features in relational data. The language is distinguished by a very simple syntax, clear and succinct semantics, and high expressivity. A wide range of feature types that have previously been utilized in statistical relational learning, probabilistic inductive logic programming, and graph mining allow for a simple and coherent representation by type extension trees. Type extension trees can be used in connection with many different model induction frameworks. In particular, we describe in this talk a kernel function on type extension trees, which allows us to exploit the associated features in the context of popular statistical learning algorithms such as support vector machines. We report results on classification experiments using some benchmark relational datasets.

Keywords: Feature construction

Joint work of: Jaeger, Manfred; Passerini, Andrea; Frasconi, Paolo

Prospects for learning causal knowledge in relational domains

David Jensen (Univ. of Massachusetts - Amherst, USA)

Recent advances in learning logical and relational models have greatly extended the range of phenomena that machine learning can address. This work has also deepened the human understanding that such models can support, largely because of their greater expressivity and interpretability. However, these advantages also encourage potential users to interpret such models causally, regardless of whether such interpretations are valid for the particular model they are examining. While researchers in AI, statistics, and philosophy have made major advances in learning causal models in the past 20 years, these advances have not been applied to learning logical and relational models. Can they be? Do logical and relational representations make causal inference easier or more difficult?

In this talk, I will briefly survey current understandings of causality, outline several reasons why researchers in PLRL should be interested in causality, and summarize recent advances in learning causal models of propositional domains. Then I will identify key areas where logical and relational representations interact with causal inference, making it either easier or more challenging. This section of the talk will utilize concrete examples, including stock fraud detection and housing mobility, and will draw on work in both machine learning, statistics, and social science. Finally, I will conclude with key research questions that define the interface of these two research areas.

Bayesian Conditioning in Infinite Domains: The Trouble with Zero

Kathryn Laskey (George Mason Univ. - Fairfax, USA)

Bayesian probability is a theory of belief dynamics. Bayesian condition on knowns, represent information about unknowns as a probability distribution, and learn by conditioning on new observations. Bayesian theory represents the universe of possibilities as a set, and specifies probabilities for subsets of the universe of possibilities. Traditional representations for the universe of possibilities have been rather simple: for example, the universe may consist of a vector of unobservable real-valued parameters and a sequence of observables that are independently and identically distributed given the parameter vector.

Classical first-order logic provides a rich language for compactly describing problems involving multiple entities interacting in complex ways. The traditional way to attach meaning to the expressions of a logical theory is to view the theory as defining a set of "possible worlds." The sentences in the language refer to entities in a set called the domain of discourse. A possible world for a theory (also called an interpretation, or model of the axioms) associates each function and relation symbol of the language to a function or relation on the domain of

discourse, in such a way that each of the axioms of the theory is a true statement about the possible world.

A dominant theme in recent proposals to integrate probability and classical logic is to develop expressive languages that allow the knowledge engineer to specify probability distributions on possible worlds. That is, the set of possible worlds defined by the axioms of a logical theory becomes the universe of possibilities on which probabilities are defined. Numerous emerging probabilistic languages provide compact representations of probability distributions for realistically complex problems. These languages use conditional independence relationships, often expressed via directed or undirected graphs with attached parameters, to exploit structure in the domain and to achieve statistical efficiency and tractability.

In finite domains (domains having finitely many entities in the domain of discourse, and finitely many function and relation symbols), the unification of probability and logic is straightforward. Sentences that follow from the axioms have probability 1; sentences that contradict the axioms have probability zero; and sentences that can be neither proven nor disproven from the axioms have probability strictly between zero and 1. But restricting to finite domains means restricting to propositional theories described in first-order syntax. The true power of both first-order logic and statistics comes in the passage to the infinite limit. But in the infinite limit, the unification of probability with logic becomes more complex. For example, each individual value in the range of a continuous random variable has probability zero, yet each is possible and one will occur. Each sequence of zeros and ones in an infinitely long series of trials of a Bernoulli random variable has probability zero, yet one of the sequences will occur. Moreover, each sequence is logically possible, yet the set of sequences with limiting frequency equal to the Bernoulli parameter has probability one. Thus, a probabilistic theory on an infinite domain may have different categories of probability zero outcomes. Some are logically impossible given the axioms of the theory; others are logically possible but not consistent with the probabilistic predictions of the theory; and still others are consistent with the probabilistic predictions of the theory.

Probability zero outcomes are important to a broad range of applications of Bayesian learning. To a Bayesian, learning is conditioning on observation. Many learning problems involve continuous distributions, so that every observation has probability zero. The result of Bayesian conditioning is arbitrary when the conditioning event has probability zero. In practice, this does not pose major difficulty, because for many commonly applied distributions, continuity assumptions yield a natural conditional distribution. I will discuss the problem of extending the continuity approach to the problem of defining conditional distributions given probability zero events in general first-order Bayesian theories.

Bayesian Logic in Infinite Domains: The Trouble with Zero

Kathryn Laskey (George Mason Univ. - Fairfax, USA)

A natural approach to integrating logic and probability is to assign probabilities to interpretations of a set of logical axioms. The logical axioms specify constraints on possible worlds, and probabilities assign likelihoods consistently to worlds that satisfy the constraints. Full integration of logic and probability requires theory and methods for treating sets of worlds that are logically possible but have probability zero. For example, statistical theories on continuous domains such as the real numbers assign probability zero to every individual outcome, yet one of the outcomes is bound to occur. For this reason, naively identifying probability zero with falsehood and probability 1 with truth is untenable. This paper presents a family of distributions for which a well-defined conditional distribution exists given any sentence consistent with the logical axioms of a theory.

Keywords: First-order logic, Bayesian reasoning, knowledge representation, random variable, Dirichlet process

Learning Relational Descriptions of Differentially Expressed Gene Groups

Nada Lavrac (Jozef Stefan Institute - Ljubljana, SLO)

Data in bioinformatics are typically multidimensional and noisy, and the phenomena to be analyzed are complicated. Precise models are often not known in advance, and must be learned from the data. This makes bioinformatics an interesting and challenging application area for machine learning. Such an interesting area is DNA microarray data analysis. A DNA microarray is a collection of microscopic DNA spots attached to a solid surface, such as glass, plastic or silicon chip forming an array. Scientists use DNA microarrays to measure the expression levels of large numbers (thousands) of genes simultaneously. Over the past few years, due to the popularization of DNA microarray technology the possibility of obtaining experimental data has significantly increased. Nevertheless, the interpretation of the results, which involves translating these data into useful biological knowledge, still remains a challenge.

This paper presents a method that uses gene ontologies, together with the paradigm of relational subgroup discovery, to find compactly described groups of genes differentially expressed in specific cancers. We applied the proposed method to three gene expression data sets with the following respective sets of sample classes: (i) acute lymphoblastic leukemia (ALL) vs. acute myeloid leukemia (AML), (ii) seven subtypes of ALL, and (iii) fourteen different types of cancers.

In our approach, the biological knowledge is composed from 4 sources of publicly available data:

1. Gene Ontology (GO) - a controlled vocabulary used to describe the biology of a gene product in any organism. There are 3 independent sets of vocabularies, or ontologies, that describe the molecular function of a gene product, the biological process in which the gene product participates, and the cellular component where the gene product can be found.
2. Kyoto Encyclopedia of Genes and Genomes (KEGG) - a collection of manually drawn pathway maps (set of genes) representing the knowledge on the molecular interaction and reaction networks for: Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes and Human Diseases.
3. Gene annotations - attached biological information to genes. This is usually done by annotating each gene with a set of GO and KEGG terms that describe its activity in the cell.
4. Gene-Gene interactions - while one gene may make only one protein, the effects of those proteins usually interact. This information is provided as gene-gene interactions.

The groups of genes differentially expressed in specific cancers are described by means of relational logic features, extracted from publicly available gene ontology information, and are straightforwardly interpretable by medical experts.

Our methodology is composed of two independent steps.

1. In the first step genes of interest are selected, in our case top K most differentially expressed genes. This is done using t-test scores of the genes. The t-test assesses whether the mean of the gene expression in one class is statistically different from the mean of the other classes.
2. In the second step we try to describe those genes in terms of the background biological knowledge. While in traditional machine learning examples are described by a tuple of values corresponding to some predefined, fixed set of attributes, a gene annotation does not straightforwardly correspond to a fixed attribute set, as it has an inherently relational character. For example, a gene may be related to a variable number of cell processes, can play a role in variable number of regulatory pathways etc. This imposes 1-to-many relations which are hard to be elegantly captured within an attribute set of a fixed size. Furthermore, a useful piece of information about a gene g may for instance be expressed by the following feature: gene g interacts with another gene whose functions include protein binding. which is elegantly captured in the form of a logical feature: interaction(g,G), function(G,protein binding).

In summary, we have approached the task of relational data mining domain by employing the methodology of relational subgroup discovery implemented in the RSD algorithm. RSD was used for the construction of relational features and for the search of subgroups of genes having common features. Using RSD we were able to discover knowledge such as:

The expression of genes coding for proteins located in the integral-to- membrane cell component, whose functions include receptor activity, has a high correlation with the BCR class of acute lymphoblastic leukemia (ALL) and a low correlation with the other classes of ALL.

Since genes frequently have multiple functions that they may be involved in, they may under some of the conditions exhibit the behavior of genes with one function and in other conditions exhibit the behavior of genes with a different function. Here subgroup discovery is effective at selecting a specific function, and in including the same gene in several subgroups.

Significant number of discovered groups of genes had a description which highlighted the underlying biological process that is responsible for distinguishing one cancer class from the other classes. The accuracy of the discovered descriptions was also verified by crossvalidation. We believe that the presented approach will significantly contribute to the application of relational data mining to gene expression analysis, given the expected increase in both the quality and quantity of gene/protein annotations in the near future.

Keywords: Relational data mining, relational subgroup discovery, bioinformatics microarray data analysis, first-order features

Joint work of: Trajkovski, Igor; Zelezny, Filip; Lavrac, Nada; Tolar, Jakub

Simple Classification Models for Networked Data: Towards combining link and attribute information

Sofus Attila Macskassy (Fetch Technologies Inc. - El Segundo, USA)

This presentation is about using multiple types of information for classification of networked data in a semi-supervised setting: given a fully described network (nodes and edges) with known labels for some of the nodes, predict the labels of the remaining nodes. I will specifically focus on the use of simple classification models such as guilt-by-association, where the models use only known class labels and known links between nodes. The primary question in such a setting is how to create these edges and how to set their weights or types. As just mentioned, one method recently developed for doing such inference is a guilt-by-association model. This method has been independently developed in two different settings–relational learning and semi-supervised learning. In relational learning, the setting assumes that the networked data has explicit links such as hyperlinks between web-pages or citations between research papers. The semi-supervised setting assumes a corpus of non-relational data and creates links based on similarity measures between the instances. Both use only the known labels in the network to predict the remaining labels but use very different information sources. The thesis of this paper is that if we combine these two types of links, the resulting network will carry more information than either type of link by itself. We test this thesis on six benchmark data sets, using a within-network learning algorithm, where we show that we gain significant improvements in predictive performance by combining the links. We describe a principled way of combining multiple types of edges with different edge-weights and semantics using an objective graph measure called node-based assortativity. I here investigate the use of this measure to combine text-mined links with explicit links and

show that using our approach significantly improves performance of our classifier over naively combining these two types of links.

(This presentation covers the material from an upcoming AAAI-2007 paper)

Generalization and consistency in structured labeling.

David McAllester (Toyota Technological Institute - Chicago, USA)

In structured labeling we are interested in input-output pairs of the form <x,y> where x and y are structured objects such as an input image and an output segmentation or an input database and an output database of "consequences". Various generalizations of support vector machines are possible for structured labeling. In this talk I will discuss structured SVMs with an emphasis on generalization bounds and consistency (optimality in the limit of infinite training data). Of particular interest is an apparent fundamental trade-off between consistency and convexity. Time permitting, some new results will be given for machines involving structured latent variables not present in the label.

Hierarchical Bayesian Modeling of Probabilistic Planning Rules

Brian Milch (MIT - Cambridge, USA)

A set of probabilistic planning rules can serve as a compact, relational model of how the world tends to change in response to an agent's actions. I will discuss our recent work on learning such rule sets for multiple related tasks – e.g., for blocks world environments with varying dynamics. We take a hierarchical Bayesian approach, in which the system learns a prior distribution over rule sets. The learned parameter of the prior distribution is a "rule set prototype" that captures commonalities across the tasks. In our generative model, this rule set prototype is modified stochastically to produce task-specific rule sets.

We have developed a coordinate ascent algorithm that iteratively optimizes the task-specific rule sets and the prior distribution. Experiments using this algorithm show that transferring information from related tasks significantly reduces the amount of training data required to predict action effects in blocks-world domains. This talk will focus not on the algorithm, but rather on the issues that arise in defining prior distributions over rule sets and other relational probabilistic models.

Keywords: Relational probabilistic models, learning, prior distributions, hierarchical bayes

Joint work of: Deshpande, Ashwin; Milch, Brian; Zettlemoyer, Luke; Kaelbling, Leslie

Learning Probabilistic Relational Dynamics for Multiple Tasks

Brian Milch (MIT - Cambridge, USA)

The ways in which an agent's actions affect the world can often be modeled compactly using a set of relational probabilistic planning rules. This extended abstract addresses the problem of learning such rule sets for multiple related tasks. We take a hierarchical Bayesian approach, in which the system learns a prior distribution over rule sets. We present a class of prior distributions parameterized by a rule set prototype that is stochastically modified to produce a task-specific rule set. We also describe a coordinate ascent algorithm that iteratively optimizes the task-specific rule sets and the prior distribution. Experiments using this algorithm show that transferring information from related tasks significantly reduces the amount of training data required to predict action effects in blocks-world domains.

Keywords: Hierarchical Bayesian models, transfer learning, multi-task learning, probabilistic planning rules

Joint work of: Deshpande, Ashwin; Milch, Brian; Zettlemoyer, Luke S.; Kaelbling, Leslie Pack

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1384

Abductive SLPs for metabolic network learning

Stephen H. Muggleton (Imperial College London, GB)

We revisit an application developed originally using Inductive Logic Programming (ILP) by replacing the underlying Logic Program (LP) description with a Stochastic Logic Program (SLP). In both the ILP and Probabilistic ILP (PILP) cases a mixture of abduction and induction are used. The PILP approach leads to a significant decrease in error accompanied by improved insight from the learned result. In (Tamaddoni-Nezhad et al, 2006) we described a logic-based representation for modelling inhibition in metabolic networks.

The approach used a variant of ILP. The example data was derived from studies of the effects of toxins on rats using Nuclear Magnetic Resonance (NMR) time-trace analysis of their biofluids together with background knowledge representing a subset of the Kyoto Encyclopedia of Genes and Genomes (KEGG).

The PILP approach applied in this paper is based on a general approach to introducing probability labels within a standard scientific experimental setting involving control and treatment data.

Keywords: PILP, SLPs, metabolic networks

Joint work of: Muggleton, Stephen H.; Chen, Jianzhong ; Santos, Jose

See also: A. Tamaddoni-Nezhad, R. Chaleil, A. Kakas, and S.H. Muggleton. Application of abductive ILP to learning metabolic network inhibition from temporal data. Machine Learning, 64:209-230, 2006. DOI: 10.1007/s10994-006-8988-x.

Exploiting prior knowledge in Intelligent Assistants - Combining relational models with hierarchies

Sriraam Natarajan (Oregon State University, USA)

Statitsical relational models have been successfully used to model static probabilistic relationships between the entities of the domain.

In this talk, we illustrate their use in a dynamic decison-theoretic setting where the task is to assist a user by inferring his intentional structure and taking appropriate assistive actions. We show that the statistical relational models can be used to succintly express the system's prior knowledge about the user's goal-subgoal structure and tune it with experience. As the system is better able to predict the user's goals, it improves the effectiveness of its assistance. We show through experiments that both the hierarchical structure of the goals and the parameter sharing facilitated by relational models significantly improve the learning speed.

Keywords: Statistical Relational Learning, Intelligent Assistants

Joint work of: Natarajan, Sriraam; Tadepalli, Prasad ; Fern, Alan

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1385

A General Importance Sampling Algorithm for Probabilistic Programs

Avi Pfeffer (Harvard University, USA)

First-order probabilistic modeling languages are capable of representing many interesting models, in a compact and elegant manner.

These models may be quite complex, so approximate inference algorithms are needed to reason about them. Many of these languages use knowledge-based model construction to construct a Bayesian network (BN) that captures the model, and then apply standard BN approximate inference algorithms to it. However, the BN itself might be very large, and once the BN is constructed much of the structure present in the model is lost.

IBAL is a highly expressive, general probabilistic modeling language that is ideal for creating new modeling frameworks. In IBAL, a model looks like a program in a functional programming language that describes the way the world is generated. We demonstrate how easy it is to create powerful new models in IBAL

using a grammar of musical transformations. IBAL also provides a rich array of program structures that make it ideal for developing inference algorithms that directly work with the structure of a program. We present an importance sampling algorithm for approximate inference for the IBAL language. Importance sampling can be quite difficult to do successfully in expressive languages because of the many deterministic relationships between concepts. Our algorithm rests on the principle of exploiting the structure in a model to infer as much as possible about a variable before committing to a value for it. It works directly on models in the language rather than constructing a BN. Our results are dramatic. We are able to obtain a very high recognition rate for our musical transformation language, whereas a rejection sampling algorithm or naive importance sampling algorithm fails miserably. We obtain similarly good performance on probabilistic context free grammars.

Keywords: First-order probabilistic languages, approximate inference

Variational Bayes via Propositionalization

Taisuke Sato (Tokyo Institute of Technology, J)

We propose a unified approach to VB (variational Bayes) in symbolic-statistical modeling via propositionalization.

By propositionalization we mean, broadly, expressing and computing probabilistic models such as BNs (Bayesian networks) and PCFGs (probabilistic context free grammars) in terms of propositional logic that considers propositional variables as binary random variables.

Our proposal is motivated by three observations. The first one is that PPC (propositionalized probability computation), i.e. probability computation formalized in a propositional setting, has turned out to be general and efficient when variable values are sparsely interdependent. Examples include (discrete) BNs, PCFGs and more generally PRISM which is a Turing complete logic programming language with EM learning ability we have been developing, and computes probabilities using graphically represented AND/OR boolean formulas. Efficiency of PPC is classically testified by the Inside-Outside algorithm in the case of PCFGs and by recent PPC approaches in the case of BNs such as the one by Darwiche et al. that exploits 0 probability and CSI (context specific independence).

Dechter et al. also revealed that PPC is a general computation scheme for BNs by their formulation of AND/OR search spaces.

Second of all, while VB has been around for sometime as a practically effective approach to Bayesian modeling, it's use is still somewhat restricted to simple models such as BNs and HMMs (hidden Markov models) though its usefulness is established through a variety of applications from model selection to prediction. On the other hand it is already proved that VB can be extended to PCFGs and is efficiently implementable using dynamic programming. Note that PCFGs

are just one class of PPC and much more general PPC is realized by PRISM. Accordingly if VB is extended to PRISM's PPC, we will obtain VB for general probabilistic models, far wider than BNs and PCFGs.

The last observation is that once VB becomes available in PRISM, it saves us a lot of time and energy. First we do not have to derive a new VB algorithm from scratch for each model and implement it. All we have to do is just to write a probabilistic model at predicate level. The rest of work will be carried out automatically in a unified manner by the PRISM system as it happens in the case of EM learning. Deriving and implementing a VB algorithm is a tedious error-prone process, and ensuring its correctness would be difficult beyond PCFGs without formal semantics.

PRISM augmented with VB will completely eliminate such needs and make it easy to explore and test new Bayesian models by helping the user cope with data sparseness and avoid over-fitting.

Keywords: Variational Bayes, propositionalized probability computation, PRISM

Joint work of: Sato, Taisuke; Kameya, Yoshitaka ; Kurihara, Kenichi

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2008/1386

Structural Sampling for Statistical Software Testing

Michele Sebag (INRIA Futurs - Orsay, F)

Structural Statistical Software Testing exploits the control flow graph of the program being tested to construct test cases.

While test cases can easily be extracted from *feasible paths* in the control flow graph, that is, paths which are actually exerted for some values of the program input, the feasible path region is a tiny fraction of the graph paths (less than 10^{-5}] for medium size programs).

The S4T algorithm presented in this paper aims to address this limitation; as an Active Relational Learning Algorithm, it uses the few feasible paths initially available to sample new feasible paths. The difficulty comes from the non-Markovian nature of the feasible path concept, due to the long-range dependencies between the nodes in the control flow graph.

Experimental validation on real-world and artificial problems demonstrates significant improvements compared to the state of the art.

Keywords: Active Relational Learning, Software Testing, Autonomic Computing, Parikh Maps

Joint work of: Baskiotis, Nicolas; Sebag, Michele

Infinite Hidden Relational Models

Volker Tresp (Siemens - München, D)

Infinite hidden relational models (IHRMs) apply nonparametric mixture modeling to relational data. An IHRM introduces for each entity an infinite-dimensional latent variable as part of a Dirichlet process (DP) mixture model, which leads to three advantages. First, IHRM reduces the extensive structural learning, which is particularly difficult in relational models due to the huge number of potential parents. Second, the information propagates globally in the ground network defined by the relationship structure. Third, the number of mixture components for each entity class can be optimized by IHRM itself based on the data. IHRM can be applied to entity clustering and relation (link) prediction, which are two important tasks in relational learning. For inference, we developed five methods: Gibbs sampling with the Chinese restaurant process, Gibbs sampling with the truncated stick breaking, and Gibbs sampling with Dirichlet-multinomial allocation, as well as two corresponding mean-field approximations. The performance of IHRM is evaluated in two domains: a recommendation system based on the MovieLens data set and prediction of the function of yeast genes/proteins on the data set of KDD Cup 2001. The experimental results demonstrates that IHRM gives significantly improved prediction about attributes of relationships and provides interpretable clustering results in complex relational models.

Full Paper:
http://wwwbrauer.informatik.tu-muenchen.de/~trespvol/papers/uai06_relation.pdf

Logical Particle Filtering

Luke Zettlemoyer (MIT - Cambridge, USA)

In this paper, we consider the problem of filtering in relational hidden Markov models.

We present a compact representation for such models and an associated logical particle filtering algorithm. Each particle contains a logical formula that describes a set of states.

The algorithm updates the formulae as new observations are received.

Since a single particle tracks many states, this filter can be more accurate than a traditional particle filter in high dimensional state spaces, as we demonstrate in experiments.

Keywords: Particle filter, logical hidden Markov model

Joint work of: Zettlemoyer, Luke; Pasula, Hanna; Pack Kaelbling, Leslie

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2008/1379