05051 Abstracts Collection
Probabilistic, Logical and Relational Learning - Towards a Synthesis
— Dagstuhl Seminar —

Luc De Raedt[1], Tom Dietterich[2], Lise Getoor[3] and Stephen H. Muggleton[4]

[1] Univ. Freiburg, DE
deraedt@informatik.uni-freiburg.de
[2] Oregon State Univ., US
tgd@cs.orst.edu
[3] Univ. of Maryland, US
getoor@cs.umd.edu
[4] Imperial College London, GB
shm@doc.ic.ac.uk

Abstract. From 30.01.05 to 04.02.05, the Dagstuhl Seminar 05051 "Probabilistic, Logical and Relational Learning - Towards a 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. Statistical relational learning, probabilistic logic learning, inductive logic programming, knowledge representation, machine learning, uncertainty in artificial intelligence

05051 Executive Summary – Probabilistic, Logical and Relational Learning Toward a Synthesis

A short report on the Dagstuhl seminar on Probabilistic, Logical and Relational Learning – Towards a Synthesis is given.

Keywords: Reasoning about Uncertainty, Relational and Logical Representations, Statistical Relational Learning, Inductive Logic Programming

Joint work of: De Raedt, Luc; Dietterich, Tom; Getoor, Lise; Muggleton, Stephen

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/412

Exploiting independence for branch operations in Bayesian learning of C&RTs

Nicos Angelopouos (University of York, GB)

In this paper we extend a methodology for Bayesian learning via MCMC, with the ability to grow arbitrarily long branches in C&RT models. We are able to do so by exploiting independence in the model construction process. The ability to grow branches rather than single nodes has been noted as desirable in the literature.

The most singular feature of the underline methodology used here in comparison to other approaches is the coupling of the prior and the proposal. The main contribution of this paper is to show how taking advantage of independence in the coupled process, can allow branch growing and swapping for proposal models.

Keywords: Bayesian machine learning, classification and regression trees, stochastic logic programs

Joint work of: Angelopoulos, Nicos; Cussens, James

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/415

Focus Problem: Natural Language Processing

Wray Buntine (Helsinki Institute for Information Technology, FIN)

This summarises the position of Natural Language Processing as a focus problem for PLRL, gives some background on problems and resources, and then contributes some problems.

Keywords: Stochastic lexicalized grammars, semantic disambiguation, structural inference in SRMs/PRMs, question answering, semantic role filling

Probabilistic Relational Learning and Natural Language

Wray Buntine (Helsinki Institute for Information Technology, FIN)

A current shortcoming in NLP is the incorporation of semantic knowledge. We can view the task of parsing a single sentence as a structural inference problem on a probabilistic relational model given by the distributional semantics of the word forms appearing in the sentence together with a grammatical model that specifies the allowed relations. For instance, "I saw the astronomer with the telescope" might yield two possible sets of relations:

verb(pronoun, "see", "astronomer") + prep-phrase("see","with","telescope")

or:
verb(pronoun, "see", "astronomer") +
prep-phrase("astronomer","with","telescope")

where other aspects of the syntax have been dealt with by the grammatical model. If a parse tree is to be built, then the task becomes one of finding a set of top-K minimum spanning trees consistent with grammatical constraints using the distributional semantics of relations as the cost function for a tree. The main problem with these kinds of relations is that the space of symbols has size 10-100 thousand, and thus standard counting methods of PRMs, tri-grams, Bayesian networks, etc., cannot be employed effectively.

Here I demonstrate how new methods for independent component analysis (ICA, MPCA, LDA, GaP, etc.) can be used to develop data-efficient probabilistic models for the relations a semantic parser would need.

Of course, one would also like to make use of any pre-existing semantic information that might be available in the form of ontologies as well. Our claim then is that probabilistic relational learning needs to incorporate these independent component methods, as well as a capability to deal with ISA concept hierarchies or other ontologies in order to provide the semantics needed for NLP.

Keywords: Independent component analysis; semantic parsing

Joint work of: Buntine, Wray; Perttu, Sami

Integrating by Separating: Combining Probability and Logic with ICL, PRISM and SLPs

James Cussens (University of York, GB)

This talk will describe the close relationship that obtains between the ICL, PRISM and SLP frameworks. The common feature of these frameworks is that a purely probabilistic component and a purely logical component are connected to produce a hybrid model. A hidden Markov model (HMM) is used as a running example. The Uniqueness Condition, which allows these frameworks to represent statistical models is discussed, and the consequences of using a weakened version of the Uniqueness Condition briefly explored. ‘Lazy’ sampling, based on SLD-resolution, is discussed.

From ILP to Probabilistic ILP

Luc De Raedt (Universität Freiburg, D)

Probabilistic inductive logic programming (sometimes also called statistical relational learning) addresses one of the central questions of artificial intelligence: the integration of Probabilistic reasoning with first order Logic representations and machine Learning.

In this talk, I shall start from an inductive logic programming perspective and sketch how it can be extended with probabilistic methods. More specifically, I shall outline three settings for inductive logic programming: learning from entailment, learning from interpretations and learning from proofs or traces and show how they can be used to learn different types of probabilistic representations.

The learning from entailment setting is natural when learning stochastic context free grammars and their upgrade, stochastic logic programs, the learning from interpretations settings is the method of choice when learning bayesian networks or bayesian logic programs, and learning from proofs or traces correspond to learning (hidden) markov models and their first order upgrades.

This work was published at ALT/DS 2004.

Joint work of: De Raedt, Luc; Kersting, Kristian

Full Paper:
http://www.informatik.uni-freiburg.de/~ml/papers/RaedtK04.pdf

See also: De Raedt, L., Kersting, K., Probabilistic Inductive Logic Programming, Proc. ALT., LNCS 3244, 2005

Markov Logic: Representation, Algorithms and Applications

Pedro Domingos (University of Washington, USA)

Markov logic adds probability to first-order logic by viewing clauses as soft constraints: the more groundings of a clause a world violates, the less probable it is. Each clause has a weight representing the strength of the corresponding constraint, and the resulting probabilistic model is a Markov random field. In this talk I will describe the algorithms we are developing for efficient learning and inference in this powerful representation. Parameters are learned by optimizing a pseudo-likelihood function using the L-BFGS algorithm. Structure learning uses this as a subroutine, and is greatly sped up by systematically avoiding redundant computation of sufficient statistics. MAP inference (finding the most probable world) is carried out using a weighted satisfiability solver, modified with a closed-world assumption to require space only on the order of the number of true ground predicates and unsatisfied ground clauses (rather than the total number of ground predicates and clauses, usually vastly larger). Probabilities are computed using a similarly modified Gibbs sampler. Discriminative learning is accomplished using a voted perceptron, with the satisfiability solver in place of the Viterbi algorithm. A few lines of Markov logic suffice to write a powerful algorithm for multi-relational object matching. After successfully applying it to the Cora dataset, we are currently using it to deduplicate a database of over half a million computer science papers.

Joint work of: Domingos, Pedro; Kok, Stanley; Parag; Richardson, Matt

Keywords: Markov random fields, pseudo-likelihood, L-BFGS, MAP inference, satisfiability solvers, Gibbs sampling, voted perceptron, object matching, deduplication

Focus Problem: Relational Reinforcement Learning

Saso Dzeroski (Jozef Stefan Institute - Ljubljana, SLO)

Reinforcement learning (RL) offers a general framework and a set of methods for constructing intelligent agents that (learn to) optimize their behavior in stochastic environments with minimal supervision.

While most work on RL is based on propositional representations, relational reinforcement learning (RRL) considers relational representations of states of the environment, agent actions, and agent policies (for choosing appropriate actions). The use of relational representations should facilitate the application of RL to complex real-world tasks, where different objects and relations among them occur naturally.

There is ample motivation to develop RRL methods, as the use of relational representations and learning methods promises many advantages. These include generalization across objects and transfer of knowledge across tasks. In addition, different kinds of prior knowledge may be used in relational learning and thus in RRL (at least in principle): e.g., background knowledge defining useful concepts and relations in the domain at hand or existing policies for solving the task at hand (that should be improved upon).

Finally, relational representations can also enable (via effective knowledge representation) the compact representation of complex and effective policies — these policies are then, by their compactness, often easier to find using search or machine learning.

The state of the art in RRL is that a number of propositional approaches to RL have been lifted to the relational case.

Q-learning was the first to be lifted, and a number of relational regression approaches to approximate the Q-function have been considered. More recently, relational approaches to approximate policy iteration, symbolic dynamic programming, and direct value-function approximation have been proposed.

However, many research issues in RRL remain wide-open for consideration. These include:

- Theory of RRL
- Hierarchical RRL
- Model Learning
- Policy Learning
- Satisficing
- Use of prior knowledge
- Bootstrapping (using guidance) in RRL
- Knowledge representation for policy and value functions
- Reasoning after solutions
- POMDP's

The aim of the present focus group is to discuss (some of) the above research issues, propose additional ones and identify the most promising and important research questions and research directions.

Currently, RRL has been illustrated on toy domains, such as the blocks-world or some computer games. Identifying killer-applications for RRL is thus an issue of crucial importance for the development of RRL. The task of identifying application areas, be it real-world or artificial, that can demonstrate the usefulness and key advantages for RRL and can serve as challenges for the development of new and improved RRL methods, will also be addressed within the focus group.

Joint work of: Driessens, Kurt; Dzeroski, Saso; Givan, Robert

Learning Logical Cost Models for Relational Sequences

Alan Fern (Oregon State University, USA)

I'll discuss my recent work in leveraging “nearly sound” logical constraints for relational sequential inference—i.e. the problem of inferring a sequence of relational states from a sequence of relational observations.

I'll introduce a simple-transition cost model, which is parameterized by weighted constraints and a state-transition cost. Inference for this model, i.e. finding a minimum-cost state sequence, reduces to a single-state minimization (SSM) problem, which becomes MAX-SAT for logical constraints. When the constraints are nearly sound and Horn, I describe a practically efficient approach to SSM based on logical inference and bounded search. I'll present a learning method that discovers constraints using the ILP system Claudien, and tunes their weights using perceptron updates. Experiments in relational video interpretation show that our learned models improve on a variety of competitors. I'll conclude by discussing the primary limitations of the framework, along with future directions and applications.

Focus Problem: Citation Analysis

Lise Getoor (University of Maryland, USA)

The first focus problem domain is bibliographic citation analysis. A large number of PLL/SRL researchers have worked with this domain.

Some advantages of this domain are: 1) the availability of data (thanks largely to Andrew McCallum, William Cohen, Steve Lawrence and others) 2) the ease of understanding the domain and 3) our obvious inherent interest in the domain as academics, :).

Within this domain, some of the objects are papers, authors, affiliations and venues and so on, and some of the links or relationships are citations, authorship and co-authorship and so on. An interesting aspect of the problem is that one must deal with indentity uncertainty: objects can be referenced in many ways, and an important task is entity resolution: figuring out the underlying object domains and mappings between references and objects.

Some of the SRL tasks in this domain include:

- topic prediction: collective classification of the topics of papers
- author attribution: predicting the author of a paper. An issue is whether we assume a closed or open world for the authors.
- Plagarism detection.
- author-topic identification: discovering the topic areas for authors. This can be used for example to assign reviewers for papers.
- entity resolution: collective clustering of the reference to objects to determine the set of authors, papers and venues.
- topic evolution: tracking change in topics over time.
- group detection: finding collaboration networks.
- citation counting/ranking: predictiing number of citations or ranking based on predicted number of citations.
- hidden object invention: Analogus to hidden variable introduction, the introduction of a hidden object, such as an advisor, that relates two author instances.
- predicate invention: from co-author information, affiliation information and perhaps information such as position and room location, invent advisor predicate.

Data sets are available:

- Many people have constructed data sets by crawling bibliography servers such as CiteSeer, ACM, DBLP and, soon one would imagine, GoogleScholar.
- Steve Lawrence several years ago made available a large collection of the citeseer data, this is available by contacting him.
- Several versions of the Cora data set are available here: http://www.cs.umass.edu/~mccallum/code-data.html

If you have additional pointers to data sets, or suggestions of additional tasks, please send email to Lise, getoor@cs.umd.edu, and we will add them to the lists.

Classification-based Relational Reinforcement Learning

Robert Givan (Purdue University, USA)

I present our thread of work on reducing reinforcement learning to cost-sensitive classification by representing and learning only policies (not value functions). Given a learned policy, learning is iterated by using policy rollout to generate training data from an improved policy. The technique is sensitive to the knowledge representation for policies, and in some domains to the bootstrapping method used to get an initial policy. When these issues are overcome, the resulting learned policies are often state-of-the-art, automatically learned, domain-specific planners, for a variety of challenging classical planning domains and their stochastic variants.

Joint work of: Givan, Robert; Fern, Alan; Yoon, Sungwook

Combining Bayesian Networks with Higher-Order Data Representations

Elias Gyftodimos (University of Bristol, GB)

This paper introduces Higher-Order Bayesian Networks, a probabilistic reasoning formalism which combines the efficient reasoning mechanisms of Bayesian Networks with the expressive power of higher-order logics.

We discuss how the proposed graphical model is used in order to define a probability distribution semantics over particular families of higher-order terms.

We give an example of the application of our method on the Mutagenesis domain, a popular dataset from the Inductive Logic Programming community, showing how we employ probabilistic inference and model learning for the construction of a probabilistic classifier based on Higher-Order Bayesian Networks.

Keywords: Probabilistic reasoning, graphical models

Joint work of: Gyftodimos, Elias; Flach, Peter

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/413

Approximate inference for relational Bayesian networks by importance sampling

Manfred Jaeger (Aalborg University, DK)

We present techniques for importance sampling from distributions defined by Relational Bayesian Networks. The methods operate directly on the abstract representation language, and therefore can be applied in situations where sampling from a standard Bayesian Network representation is infeasible. We describe experimental results from using standard, adaptive and backward sampling strategies. Furthermore, we use in our experiments a model that illustrates a fully general way of translating the recent framework of Markov Logic Networks into Relational Bayesian Networks.

Keywords: Relational models, Importance Sampling

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/411

Focus Problem(s): Fundamental representation tasks

Manfred Jaeger (Aalborg University, DK)

The following modeling and inference problems are meant to elucidate basic representational capabilities that probabilistic/logical representation languages may need to possess to address a variety of real application problems. All problems are stated as a representation problem for a concrete small example that is illustrative of a fundamental representation task. Goal is to compare different representation languages on how effectively they can handle these tasks. Problems are stated as modeling and inference problems, but obviously to each can be added the problem of learning such models.

Problem 1: random domain functions. Represent random functions bet - ween objects of a domain.
Example (following Getoor, Friedman, Koller & Taskar, ICML-01) a domain consists of objects of type 'movie' and 'theater'.
Movies have an attribute 'genre' with possible values 'foreign' and 'thriller'. Theaters have an attribute 'type' with possible values 'megaplex' and 'art theater'. Represent a model in which theaters randomly select a movie for showing. The probability of which genre of movie is shown depends on the type of the theater.
Given the selected genre, a movie is uniformly chosen from all movies of that genre.

Problem 2: domain uncertainty. Represent models where the number of objects in the domain is uncertain (and potentially unbounded)
Example: represent a domain consisting of all persons some patient was in contact with. Let the size k of the domain be given by e.g. a geometric or

a Poisson distribution. Let all persons in the domain be 'infected' with a certain probability, and transmit the infection to the patient with a certain probability.

Simplified version: assume that there is a maximal number of 500 people in the domain, and let the probability of k=500 be the tail probability P(k>=500) for the geometric or Poisson P().

Supported inferences e.g.: given that the domain has at least 10 people, what is the probability that the patient is infected?

Problem 3: models of first-order sentences. Represent the uniform distribution on all models (of a given size) for some first-order sentence ϕ.

Example: Let ϕ be the first-order sentence that says that the binary 'edge' relation defines an undirected graph (i.e. edge is symmetric and has no self-loops), and that the nodes of the graph are colored with the three colors 'blue', 'yellow' and 'red' such that no two connected nodes have the same color.

Represent the uniform distribution over all models of ϕ over a fixed domain of nodes.

Supported inferences e.g.: given that there is an edge between nodes a and b, and nodes b and c, and that a is yellow and b is blue, what is the probability that c is red?

Problem 4: probabilistic context-free grammars Encode the distribution defined by a probabilistic context-free grammars and support the usual inferences.

Example (taken from www.coli.uni-sb.de/ ~ schulte/Teaching/MG3-04/Vorlesung/matt_pcfg1.pdf):
Encode the pcfg:

```
S -> NP VP   1.0
PP-> P NP    1.0
VP-> V NP    0.7
VP-> VP PP   0.3
P -> with    1.0
V -> saw     1.0
NP-> NP PP   0.4
NP-> astronomers 0.1
NP-> ears    0.18
NP-> saw     0.04
NP-> stars   0.18
NP-> telescopes  0.1
```

Support inferences: probabilities of words, most probable parse-trees for a given word.

Simplified version: restrict model to words or productions of a fixed maximal length.

Does accurate statistical inference require joint models of attributes and relations?

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

Recent work has produced several types of probabilistic models of relational data, including models of attributes (e.g., topic(web-page-x)) and models of links and groups (e.g., author-of(person-x, paper-y) or member-of(person-x, person-y, person-z, research-group-a)). To date, however, nearly all probabilistic models have focused on either predicting attributes conditioned on relational structure or vice versa, rather than producing a joint model of structure and attributes. Probabilistic dependencies often span the boundary between attributes and structure, and several of recent papers have demonstrated how statistical inference can be biased if these dependencies are ignored. It appears that accurate statistical inference in relational data may virtually require a joint model of attributes and structure. Without such a joint model, it is difficult to adjust for the effects of structure on attribute dependence (and vice versa), and hypothesis tests and parameter estimates are likely to be biased. This identifies a important new research challenge for work in learning probabilistic models of relational data.

Revealing information structures in data

Gabriele Kern-Isberner (Universität Dortmund, D)

One of the principal aims of knowledge discovery is to find generic, predictive patterns in data, i.e. to extract structural from numerical, statistical information. A popular approach here is to take patterns to be represented by (association) rules, and to look for rules with a high support and a high confidence (conditional probability), assuming that relevant information is somehow revealed by things that occur often.

However, generic knowledge is rather a qualitative notion, and still the question remains unsolved, how exactly structural links are reflected by statistical data. When searching for Bayesian network structures in data, the opposite problem is dealt with, since Bayesian networks are based on conditional independencies reflecting missing structural links.

This talk will focus on the discovery of structural conditional dependencies (rules) in data. The approach has its roots in default logic and probabilistic logic, considering knowledge discovery as an operation which reverses inductively completing knowledge. So, the set of rules which are extracted from data may serve as a knowledge base for further inference.

This kind of inference may be quite complex. Indeed, the method to be presented has been developed to compute probabilistic rules from data which can be used for inference on maximum entropy (ME). This principle from information theory makes sure that incomplete probabilistic knowledge is processed in a most unbiased way. The basic philosphy of our approach is to exploit algebraic structures from frequency distributions which are imposed on data when ME-learning rules, or, to put it differently, to find the footprints of the ME-principle in numerical information. In particular, it is possible to disentangle highly complex interactions of rules. As information theory is involved, the benefits of our method are twofold: First, the computed rules can be used directly for automatic ME-inference. Here, it is possible to abstract from zero probabilities occurring in data and to use information structures to derive hypothetic knowledge that has not been observed. Second, these rules can be considered as being most informative for the user, as they emerge from focusing on information theoretically relevant structures.

Keywords: Association rules, information theory, maximum entropy, inductive knowledge representation

Probabilistic Logic Reasoning over Time

Kristian Kersting (Universität Freiburg, D)

Most probabilistic logic learning approaches developed in the last years do not address sequential data in a principled way. Many real world sequences, however, can elegantly be represented as sequences of logical atoms. This talk will overview our recent work on probabilistic models for logical sequences.

More precisely, the talk will review "logical hidden Markov models" (LOHMMs). Logical hidden Markov models (LOHMMs) are a generalization of hidden Markov models (HMMs) to analyze sequences of logical atoms. In LOHMMs, abstract states summarize sets of states and are represented by logical atoms. Transitions are defined between abstract states to summarize sets of transitions between states. Unification is used to share information among states, and between states and observations. Due to logical abstraction, a LOHMM can be designed to be smaller than an equivalent HMM by an order of magnitude in the number of parameters.

Further issues addressed will be: + adaptations of classical HMM procedures such as forward-backward procedure, + results on expressivity + discriminative learning + bioinformatics applications: protein fold and mRNA signalstructure classification + connections to the currently emerging area of "relational reinforcement learning" (RRL).

Joint work of: Kersting, Kristian; De Raedt, Luc; Raiko, Tapani; Gaertner, Thomas

Learning to Act in First Order MDPs

Roni Khardon (Tufts University, USA)

The problems of learning and acting in First Order MDPs can be seen as generalizing both classical planning and reinforcement learning.

These problems have received some attention recently by several authors. Since these problems are hard, the learner is typically "helped" in some way, e.g. by using small or easy instances, and this can be seen as some form of supervision for the learner. The talk will review some theoretical and empirical results on supervised learning in this context, discuss their scope and limitations, as well as implications for future work.

Knowledge Based Operations on Graphical Models

Rudolf Kruse (Universität Magdeburg, D)

The presentation refers to new theoretical and algorithmic results on graphical models as well as some details on industrial applications.

Keywords: Probabilistic Graphical Models, Learning, Revision, Updating, Constraints

Joint work of: Kruse, Rudolf; Gebhardt, Jörg

Naive Bayesian classification of structured data

Nicolas Lachiche (University of Strasbourg, F)

We would like to present 1BC and 1BC2, two systems that perform naive Bayesian classification of structured individuals.

The approach of 1BC is to project the individuals along first-order features. These features are built from the individual using structural predicates referring to related objects (e.g. atoms within molecules), and properties applying to the individual or one or several of its related objects (e.g. a bond between two atoms). We describe an individual in terms of elementary features consisting of zero or more structural predicates and one property; these features are treated as conditionally independent in the spirit of the naive Bayes assumption. 1BC2 represents an alternative first-order upgrade to the naive Bayesian classifier by considering probability distributions over structured objects (e.g., a molecule as a set of atoms), and estimating those distributions from the probabilities of its elements (which are assumed to be independent).

Joint work of: Lachiche, Nicolas; Flach, Peter

Handling Uncertainty for Adaptive Logical Agents

John Lloyd (Australian National University - Canberra, AU)

This talk shows how Bayesian networks and decision theory can be used by adaptive logical agents to select actions.

The agents have belief bases that are theories in a multi-modal, higher-order logic. Belief bases can be modified using symbolic, on-line learning processes. A method of partitioning the state space of the agent in two different ways leads to a Bayesian network and associated influence diagram for selecting actions.

The resulting agent architecture exhibits a tight integration between logic and Bayesian networks. This approach to agent architecture is illustrated by a user agent that is able to personalise its behaviour according to the user's interests and preferences.

Joint work of: Lloyd, John; Sears, Tim

An Architecture for Rational Agents

John Lloyd (Australian National University - Canberra, AU)

This paper is concerned with designing architectures for rational agents.

In the proposed architecture, agents have belief bases that are theories in a multi-modal, higher-order logic.

Belief bases can be modified by a belief acquisition algorithm that includes both symbolic, on-line learning and conventional knowledge base update as special cases.

A method of partitioning the state space of the agent in two different ways leads to a Bayesian network and associated influence diagram for selecting actions.

The resulting agent architecture exhibits a tight integration between logic, probability, and learning.

This approach to agent architecture is illustrated by a user agent that is able to personalise its behaviour according to the user's interests and preferences.

Keywords: Rational agent, agent architecture, belief base, Bayesian networks

Joint work of: Lloyd, John; Sears, Tim

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/419

Multi-Relational Regression in Spatial Domains

Donato Malerba (University of Bari, I)

Statistical modelling of spatial data raises several distinctive problems. We consider the special case in which a regression model on area referenced data has to be built and the response variable is associated to an area. The observation for an area can be descriptive of one or more primary units, possibly of different type, within the area. In addition to attributes that relate to primary units or areas, there are attributes that refer to relationships between primary units (e.g. contact frequencies between households) and between areal units (e.g., migration rates). This relational information is often responsible for the spatial variation and it is extremely useful in modelling. When the spatial heterogeneity of response can be anticipated, the data analyst may allow either the constant or one (or more) of the other regression parameters to vary spatially, thanks to the introduction of dummy variables. In general, however, it is necessary to resort to automated methods that look for regional segmentation according to spatial heterogeneity. We report of a model tree induction method that can capture both (spatially) local and global effects of explanatory variables. We also present a multi-relational extension of the method in order to build regression models with spatially lagged explanatory variables, especially useful when the effect of an explanatory variable at any area is not limited to the specified area (e.g., the proportion of people suffering from respiratory diseases in a site also depends on the high/low level of pollution of sites where people daily move). Finally, we discuss the applicability of the multi-relational model tree induction method also to the case of spatially lagged response variables, that is, when autocorrelation affects the regressor values (e.g., the price for a good at a retail outlet in a city may depend on the price of the same good sold by local competitors).

Keywords: Spatial data mining, multi-relational regression

Joint work of: Malerba, Donato; Appice, Annalisa; Ceci, Michelangelo

A* Instead of Dynamic Programming

David McAllester (Toyota Technological Institute - Chicago, USA)

A wide variety of dynamic programming algorithms can be defined by bottom-up logic programming inference rules. Examples includes statistical context free parsing and optimal matching of deormable models in vision. We give a version of A* with patterns that can be applied to arbitrary dynamic programming algorithms formulated in this way. This allows the mechanical synthesis of a large number of novel A* algorithms. Formulating allgorithms as inference rules also allows for effective integration of different levels of processing without the need for n-best lists to be passed between levels.

BLOG: Probabilistic Models with Unknown Objects

Brian Milch (Univ. California - Berkeley, USA)

We introduce BLOG, a formal language for defining probability models with unknown objects and identity uncertainty. A BLOG model describes a generative process in which some steps add objects to the world, and others determine attributes and relations on these objects. Subject to certain acyclicity constraints, a BLOG model specifies a unique probability distribution over first-order model structures that can contain varying and unbounded numbers of objects. Furthermore, inference algorithms exist for a large class of BLOG models.

Keywords: Knowledge representation, probability, first-order logic, identity uncertainty, unknown objects

Joint work of: Milch, Brian; Marthi, Bhaskara; Russell, Stuart; Sontag, David; Ong, Daniel L.; Kolobov, Andrey

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/416

Statistical Relational Learning for Natural-Language Information Extraction and Semantic Parsing

Ray Mooney (Univ. of Texas at Austin, USA)

We review our work on using statistical relational learning (SRL) for natural-language processing (NLP), and promote NLP as an ideal application area for SRL. First, we discuss the properties of NLP problems that naturally require SRL. Next, we discuss our recent work on using Relational Markov Networks (RMNs) for information extraction (IE). Most IE systems treat separate potential extractions as independent; however, in many cases, considering influences between potential extractions could improve accuracy. By using RMNs, we can exploit arbitrary dependencies between extractions, allowing for "collective information extraction." Experiments on learning to extract protein names from biomedical text demonstrate the advantages of this approach. Finally, we discuss some older work on semantic parsing which uses committees of relational hypotheses induced by a traditional ILP system to estimate probabilities in a statistical parser. Experiments on learning to map English questions into formal database queries demonstrate the advantages of this approach.

A Comparison of the Expressive Power of some Probabilistic-logical Models

Stephen H. Muggleton (Imperial College London, GB)

Probabilistic-Logical Models (PLMs) are recognized as efficient frameworks to represent real-world problems: they combine the expressive power of first-order logic, which serves as a knowledge representation language, and the capability to model uncertainty with probabilities. Among existing models, it is usual to distinguish the domain-frequency approach from the possible-worlds approach.

Stochastic Logic Programs (SLPs) and Probabilistic Relational Models (PRMs), which are considered as domain-frequency approaches, and on the other hand Bayesian Logic Programs (BLPs) and Stochastic Relational Models (SRMs) (possible-worlds approaches), are promising PLMs in their categories.

This presentation is aimed at comparing the respective expressive power of these frameworks; we demonstrate relations between their semantics.

We identify a subclass of BLPs that can be encoded by SLPs, and lift this result to any BLP by introducing extended SLPs.

Converse properties are reviewed, and we show how BLPs can define the same semantics as complete, range-restricted SLPs. We further demonstrate that BLPs (respectively SLPs) can naturally encode the relational semantics of PRMs (respectively SRMs). Whenever applicable, we provide inter-translation algorithms, prove their soundness and give worked examples.

Keywords: Proabilistic logics, expressive power

Joint work of: Muggleton, Stephen H.; Chen, Jianzhong; Puech A.; Grisel, O.

Focus Problem: BioNetwork

Stephen H. Muggleton (Imperial College London, GB)

This focus problem is based on using machine learning to complete bio-molecular pathway descriptions. Systems biologists use graph-based descriptions of bio-molecular interactions which describe cellular activities such as gene regulation, metabolism and transcription. Biologists build and maintain these network models based on the results of experiments in wild-life and mutated organisms.

Recently an increasing number of PLL/SRL researchers have worked within this domain. Advantages of the domain include 1) the availability of background knowledge on existing known biochemical networks from publicly available resources such as KEGG (used in data sets such as those in the Nature paper by Bryant, King, Muggleton, etc); 2) the availability of training and test data from a variety of sources including micro-array experiments (see for instance the Rosetta compendium of Hughes et al.) and metabonomic data (eg Nicholson et al.) from NMR and mass spectroscopy experiments; 3) the inherent importance of the problem (see Kitano's articles in Nature and Science in 2002) owing to its application in biology and medicine; 4) the inherent relational structure in the form of spatial and temporal interactions of the molecules involved; 5) the naturalness of probabilistic representations to represent, for instance, the availability of genes as discrete or continuous random variables having expression levels as states (used by Friedman et al with Bayes' nets and Angelopoulos with Stochastic Logic Programs).

The objects within this domain include genes, proteins, metabolites, inhbitors and cofactors. The relationships include biochemical reactions in which one set of metabolites is transformed to another mediated by the involvement of an enzyme. Within various databases the same object can be referred to in several ways, which brings in the problem of identity uncertainty. The available genomic information is also very incomplete concerning the functions and even the existence of genes and metabolites, leading to the necessity of techniques such as logical abduction to introduce novel functions and even invention of new objects. This process needs to be constrained by existing chemical and biological knowledge. For instance, general laws concerning conservation of matter rule out biochemicals which cannot be constructed from the available constituents. Novel enzymic functions can often be speculated on the basis of perturbing known functions. Complete metabolic subnetworks can be hypothesised on the basis of known networks in species which are closely related by evolution.

Some of the PLL tasks in this domain include:

- function prediction: given a partial network together with expression data assign an enzyme function to a node in the network
- function class learning: given a class of functions together with a proposed new function, generalise the class to include the new function
- hidden object invention: given a partial network together with expression data assign suggest the existence of unknown intermediate metabolites
- reaction rate learning: given a partial network together with expression data assign reaction rates expressed as probabilities of their occurrences
- inhibitor prediction: given metabonomic data showing the time-varying behaviour of metabolites following the introduction of a toxin predict which reactions are being inhibited

Data sets available:

- Background knowledge for this task is available from a large, and growing set of publicly available resources. These include KEGG (biochemical networks), Brenda (inhbitors and cofactors), Ensembl (genes), Ligand (compendium of biochemical reactions) and Go (hierarchical classification of protein functions).
- Datasets from microarray experiments can be obtained from sources such as Rosetta as well as sets developed for machine learning purposes by authors such as Angelopoulos, Bryant, King, Watanabe Tamaddoni-Nezhad and Muggleton.

If you have additional pointers to data sets, or suggestions of additional tasks, please send email to Stephen Muggleton, shm@doc.ic.ac.uk, and we will add them to the lists.

Leveraging relational autocorrelation with latent group models

Jennifer Neville (Univ. of Massachusetts - Amherst, USA)

The presence of autocorrelation provides strong motivation for using relational techniques for learning and inference. Autocorrelation is a statistical dependency between the values of the same variable on related entities and is a nearly ubiquitous characteristic of relational data sets. Recent research has explored the use of collective inference techniques to exploit this phenomenon. These techniques achieve significant performance gains by modeling observed correlations among class labels of related instances, but the models fail to capture a frequent cause of autocorrelation—the presence of underlying groups that influence the attributes on a set of entities. We propose a latent group model (LGM) for relational data, which discovers and exploits the hidden structures responsible for the observed autocorrelation among class labels. Modeling the latent group structure improves model performance, increases inference efficiency, and enhances our understanding of the datasets. We evaluate performance on three relational classification tasks and show that LGM outperforms models that ignore latent group structure when there is little known information with which to seed inference.

Keywords: Statistical relational learning, probabilistic relational models, latent variable models, autocorrelation, collective inference

Joint work of: Neville, Jennifer; Jensen, David

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/420

Focus Problem: Temporal Reasoning in Relational Domains

Jennifer Neville (Univ. of Massachusetts - Amherst, USA)

Many relational datasets have a temporal aspect that is inherent to the domain. For example, in scientific publications, citation patterns, researchers' careers, and topics, all change over time. In these domains, it is challenging but necessary to incorporate temporal dynamics in order to make relational models competitive with hand-crafted propositionalized rules. Straightforward approaches, that include time as an additional modeling dimension, have shown promise but these approaches increase the dimensionality of the model and feature space to an even more intractable level than conventional relational models. More creative approaches are needed to push these solutions to the next level.

Tasks

- Predict static attribute (e.g., paper topics) use temporal dynamics of relations e.g., model citation patterns of references
- Predict dynamic attribute (e.g., number of citations 'next year' per paper) use temporal dynamics of relations and attributes e.g., model number of citations 'last year'
- Predict link/relations (e.g., paper publication venue) use temporal dynamics of relations and attributes e.g., model topics of papers published in the venue 'last year'
  - relations may appear over time (e.g., publication), or
  - relations may appear and disappear over time (e.g., author-at-institution)
- Predict emergent group behavior (e.g., new research group formation) use temporal dynamics of relations and attributes e.g., intra-group citation patterns over the 'last 3 years'

Representation Challenges

- Temporal attributes, attributes with values that change over time
- Temporal relations, events/objects that occur over time
- Non-link relations, events that occur "closely" in time are related

Modeling Issues

- Sampling: Construct samples with respect to time
- Dependencies to model:
  - Between attribute values at different time steps
  - Between attribute values and graph dynamics
  - Between attribute values that are "unrelated" but occur closely in time
  - Temporal autocorrelation
- Feature construction
  - Compare examples from different time slices
  - Temporal aggregation: Consider aggregates of attributes over time windows and consider aggregates of graph dynamics given time windows

Questions

- Is time just another type of relation? If not, what makes time special? Is space similar?
- Are there classes of problems for which temporal analysis is easy? If so, what aspects differentiate these problems?
- What are the bottleneck issues to incorporating time into relational models? How can domain knowledge be used to simplify the inclusion of time in the models? What can techniques can we use from time series analysis?

Joint work of: Neville, Jennifer; Jensen, David

View Learning for SRL, with an Application to Mammography

C. David Page (Univ. Wisconsin - Madison, USA)

This talk will use mammography and several other biomedical applications to motivate "view learning." Current SRL systems provide the ability to integrate data from multiple relational tables, and from multiple rows of the same table. SRL systems with view learning will provide the added ability to create new fields or tables – a new view of the database – in order to enhance predictive accuracy and insight. We will present an initial evaluation of view learning on the task of predicting whether abnormalities on a mammogram are benign or malignant.

Keywords: Databases, views, biomedical applications, mammography

Joint work of: Page, C. David; Davis, Jesse; Burnside, Beth; Santos Costa, Vitor; Dutra, Ines; Shavlik, Jude; Ramakrishnan, Raghu

Kernels on Prolog Proof Trees:Statistical Learning in the ILP Setting

Andrea Passerini (University of Firenze, I)

An example-trace is a sequence of steps taken by a program on a given example input. Different approaches exist in order to exploit example-traces for learning, all explicitly inferring a target program from positive and negative traces.

We generalize such idea by developing similarity measures betweeen traces in order to learn to discriminate between positive and negative ones. This allows to combine the expressiveness of inductive logic programming in representing knowledge to the statistical properties of kernel machines. Logic programs will be used to generate proofs of given visitor programs which exploit the available background knowledge, while kernel machines will be employed to learn from such proofs.

Keywords: Proof Trees, Logic Kernels, Learning from Traces

Joint work of: Passerini, Andrea; Frasconi, Paolo

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2006/417

Distribution-Based Aggregation for Relational Learning from Identifier Attributes

Claudia Perlich (New York University, USA)

Feature construction through aggregation plays an essential role in modeling relational domains with one-to-many relationships between tables. One-to-many relationships lead to bags (multisets) of related entites, from which predictive information must be captured.

This work focuses on the aggregation of categorical attributes with many possible values, in particular identifiers (e.g., keys attributes), which typically are not considered for aggregation in predictive modeling. The presented aggregation method uses meta-data about the class-conditional distributions and can be interpreted as a "relational fixed-effect" model within a Bayesian framework. We also analyze how the aggregation of key attributes allows the circumvention of limitations caused either by missing/unobserved object properties or by independence assumptions.

The empirical work is executed using the relational learning system ACORA, that allows the comparison of various aggregation methods following a propositionalization approach.

Joint work of: Perlich, Claudia; Provost, Foster

Combining logic and probability: a knowledge representation perspective

David Poole (University of British Columbia - Vancouver, CDN)

This talk will give an overview of what we want from a knowledge representation, and how a mix of probability and logic could serve as such a representation. I will survey the possible approaches, giving particular motivation for the independent choice logic (ICL). I will show some pitfalls that you need to be careful with when designing a mix of probability and logic. I will also make outrageous claims in order to create discussion, such as if you do reasoning right, you don't need anything special to do learning (you just condition on the data).

Keywords: Logic, probability, Bayesian decision theory, Bayesian networks, independent choice logic, independence

Classification in Networked Data: A toolkit and a univariate case study

Foster Provost (New York University, USA)

I will present NetKit, a modular toolkit for classification in networked data, and a case-study of its application to a collection of networked data sets used in prior machine learning research.

Networked data are relational data where entities are interconnected, and this paper considers the common case where entities whose labels are to be estimated are linked to entities for which the label is known. NetKit is based on a three-component framework, comprising a local classifier, a relational classifier, and a collective inference procedure. Various existing relational learning algorithms can be instantiated with appropriate choices for these three components and new relational learning algorithms can be composed by new combinations of components. The case study demonstrates how the toolkit facilitates comparison of different methods (which for networked data has been lacking in machine learning research). The case study focuses on the simple but important special case of univariate network classification, for which the only information available is the structure of class linkage in the network (i.e., only links and some class labels are available).

Joint work of: Provost, Foster; Macskassy, Sofus

Knowledge Representations, Learning and Inference for Natural Language Understanding

Dan Roth (Univ. of Illinois - Urbana, USA)

Natural language decisions require the use of deeper structural, relational and semantic properties of the text. There needs to be a unified knowledge representation of the text, that (1) provides a hierarchical encoding of the structural, relational and semantic properties of the given text, (2) is integrated with learning mechanisms that can be used to induce such information from newly observed raw text, and (3) that is equipped with an inferential mechanism that can be used to support inferences with respect to such complex representations.

I will describe our research on all three aspects of the problem – from the design of relational input representations, through learning approaches that exploit these structures and the relations between observations, to progress on inducing complex and structured output representations.

Depending on the time available I will focus on one or several aspects of these, and provide examples from our work on tasks such as Entity Identification, Semantic Parsing and open-domain Question Answering.

Learning through failure

Taisuke Sato (Tokyo Institute of Technology, J)

PRISM, a symbolic-statistical modeling language we have been developing since '97, recently incorporated a program transformation technique to handle failure in generative modeling.

I'll show this feature opens a way to new breeds of symbolic models, including EM learning from negative observations, constrained HMMs and finite PCFGs.

Keywords: Program transformation, failure, generative modeling

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/418

Joint work of: Sato, Taisuke; Kameya, Yoshitaka

Multi-View EM and Link Farm Discovery

Tobias Scheffer (HU Berlin, D)

The first part of this abstract focuses on estimation of mixture models for problems in which multiple views of the instances are available. Examples of this setting include clustering web pages or research papers that have intrinsic (text) and extrinsic (references) attributes. Mixture model estimation is a key problem for both semi-supervised and unsupervised learning. An appropriate optimization criterion quantifies the likelihood and the consensus among models in the individual views; maximizing this consensus minimizes a bound on the risk of assigning an instance to an incorrect mixture component. An EM algorithm maximizes this criterion. The second part of this abstract focuses on the problem of identifying link spam. Search engine optimizers inflate the page rank of a target site by spinning an artificial web for the sole purpose of providing inbound links to the target. Discriminating natural from artificial web sites is a difficult multi-view problem.

Keywords: Multi-view learning

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2006/414

Statistics in Relational Data Mining

Arno Siebes (Utrecht University, NL)

One of the popular approaches to relational mining is to use aggregate functions over 1-n relations. This approach yields promising results, but it does not automatically generalise to m-n relations, e.g., because it is not clear that an aggregate such as Average amount on bank accounts has the same meaning for all persons in a table.

Recently I have been working on more general approaches to exploit the information in the associated tables. In this talk I will explain the problems and give some examples on how one can handle the related information. It is still work in progress, so while the goal is m-n relations, the focus is still on 1-n relations

Keywords: Relational mining, aggregates, statistics

Learning with Combining Rules in a Relational Conditional Influence Language

Prasad Tadepalli (Oregon State University, USA)

The Knowledge Intensive Learning group at Oregon State University is experimenting with a probabilistic relational language for expressing prior knowledge for learning in knowledge-rich domains.

The core construct of this language is a "conditional influence" statement, which specifies the conditions under which some random variables can influence another. The conditions are expressed in terms of the known relations between different variables.

When these relationships are not one-to-one, one needs combining rules or aggregators to summarize the probability distributions. In this talk we motivate the conditional influence language, show its relationship to other formalisms, and describe some preliminary results on gradient-descent learning in the presence of combining rules.

Keywords: Conditional influences, probabilistic relational learning, combining rules, gradient descent learning, knowledge intensive learning

Joint work of: Tadepalli, Prasad; Altendorf, Eric; Dietterich, Thomas; Fern, Alan; Natarajan, Sriraam; Restificar, Angelo

Dirichlet Enhanced Probabilistic Relational Models

Volker Tresp (Siemens - München, D)

A hierarchical Bayesian (HB) approach can significantly increase the modelling power of probabilistic relational models (PRMs).

Whereas in most PRM approaches, parameters specifying conditional probabilities are global, a HB approach allows parameters to be owned by entities. This significantly increases the flexibility of PRM models; at the same time the parameters are constraint to be generated from a common prior distribution. By adapting the hyperparameters of the prior distribution, parameters for different entities can learn from one another, i.e., they can "borrow strength". In a HB model, the parametric form of the prior distribution is of greater importance than in non-hierarchical models. Since the "learned" prior distribution might not be well approximated by a standard distribution, it makes sense assume a nonparametric distribution, e.g. the prior is assumed to be generated from a Dirichlet process.

I will first introduce HB modelling and motivate a nonparametric hierarchical Bayesian modelling approach. I will discuss the properties of Dirichlet processes, which are the basis for most nonparametric Bayesian approaches. I will briefly illustrate various learning and inference approaches for Dirichlet processes and discuss applications of nonparametric HB modelling to PRM models.

Keywords: Probabilistic relational models hierarchical Bayesian modelling Dirichlet processes nonparametric Bayesian modelling

Joint work of: Tresp, Volker; Xu, Zhao

Logic Programming meets Dynamic Programming

Martijn van Otterlo (University of Twente, NL)

In this talk we present a general scheme for relational dynamic programming in relational domains, based on the previously introduced Relational Bellman Backup Operator: ReBel. We identify a number of necessary operations in this general scheme and we show how we can make use of standard techniques from (inductive) logic programming to do computations much more efficiently. This will enable solving larger problems.

In particular, the use of tabling makes it possible to store already computed abstractions. The use of tabling also highlights an important distinction between relational dynamic programming – in which new abstractions are formed – and standard dynamic programming – in which the abstraction level is fixed and values are propagated at this level.

A general question is raised about how much of either dynamic programming techniques or logic programming techniques can be used or are even necessary for relational dynamic programming.

Keywords: Relational reinforcement learning

Joint work of: van Otterlo, Martijn; Kersting, Kristian; Fischer, Joerg; De Raedt, Luc