1 Introduction

Lise Getoor and Ben Taskar

We outline the major themes, problems and approaches that define the subject of the book: statistical relational learning. While the problems of statistical learning and relational representation and reasoning have a fairly long history on their own in Artificial Intelligence research, the synthesis of the approaches is currently a burgeoning field. We briefly sketch the background and the recent developments presented in the book.

1.1 Overview

The vast majority of statistical learning literature assumes the data is represented by points in a high-dimensional space. For any particular isolated task, such as learning to detect a face in an image or classify an email message as spam or not, we can usually construct the relevant low-level features (e.g., pixels, filters, words, URLs) and solve the problem using standard tools for the vector representation. While extremely useful for development of elegant and general algorithms and analysis, this abstraction hides the rich logical structure of the underlying data that is crucial for solving more general and complex problems. We may like to detect not only a face in an image but to recognize that, for example, it is the face of a tall woman who is spiking a volleyball or a little boy jumping into a puddle, etc. Or, in the case of email, we might want to detect that an email message is not only not-spam but is a request from our supervisor to meet tomorrow with three colleagues or an invitation to the downstairs neighbor's birthday party next Sunday, etc. We are ultimately interested in not just answering an isolated yes/no question, but in producing and manipulating structured representations of the data, involving objects described by attributes and participating in relationships, actions, and events. The challenge is to develop formalisms, models, and algorithms that enable effective and robust reasoning about this type of object-relational structure of the data.

Dealing with real data, like images and text, inevitably requires the ability to handle the uncertainty that arises from noise and incomplete information (e.g., occlusions, misspellings). In relational problems, uncertainty arises on many levels. Beyond uncertainty about the attributes of an object, there may be uncertainty about an object's type, the number of objects, and the identity of an object (what kind, which, and how many entities are depicted or written about), as well as relationship membership, type, and number (which entities are related, how, and how many times). Solving interesting relational learning tasks robustly requires sophisticated treatment of uncertainty at these multiple levels of representation.

In this book, we present the growing body of work on a variety of statistical models that target relational learning tasks. The goal of these representations is to express probabilistic models in a compact and intuitive way that reflects the relational structure of the domain and, ideally, supports efficient learning and inference. The majority of these models are based on combinations of graphical models, probabilistic grammars, and logical formulae.

1.2 Brief History of Relational Learning

Early work on machine learning often focused on learning deterministic logical concepts. Methods were typically noise and mostly applied to ”toy” domains. One of the earliest relational learning systems is Winston's arch learning system [49]. This online-style system was trained using a sequence of instances labeled as positive and negative examples of arches. The system maintained a ”current” hypothesis, represented as a semantic network. When a new example was presented, the system made a prediction using the current hypothesis. If the prediction was correct, no changes were made to the hypothesis. If it was incorrect, then the set of differences between the current hypothesis and the example was identified. If the example was a positive instance, the differences were used to generalize the concept; if the example was a negative instance, it was used to specialize the concept. Following this there were a number of more advanced relational learning systems [8, 18, 45], but all used a similar logic-based representation for the concepts.

This approach of machine learning (ML) fell out of vogue for many years because of problems handling noise and large-scale data. During that time, the ML community shifted attention to statistical methods that ignored relational aspects of the data (e.g., neural networks, decision trees, and generalized linear models). These methods led to major boosts in accuracy in many problems in low-level vision and natural language processing [11, 28]. However, their focus was on the propositional or attribute-value representation.

The major exception has been the inductive logic programming (ILP) community. The ILP community has concentrated its efforts on learning (deterministic) first-order rules from relational data [27, 30]. Initially the ILP community focused its attention solely on the task of program synthesis from examples and background knowledge. However, recent research has tackled the discovery of useful rules from larger databases [9]. These rules are often used for prediction and may have a probabilistic interpretation. The ILP community has had successes in a number of application areas including discovery of 2D structural alerts for mutagenicity/carcinogenicity [22], 3D pharmacophore discovery for drug design [10], and analysis of chemical databases [7].

1.3 Emerging Trends

Recently, both the ILP community and the statistical ML community have begun to incorporate aspects of the complementary technology. Many ILP researchers are developing stochastic and probabilistic representations and algorithms [31, 21, 6]. In more traditional ML circles, researchers who have in the past focused on attribute-value or propositional learning algorithms are exploring methods for incorporating relational information [5, 32, 4]. It is our hope that this trend will continue, and that the work presented in this book will provide a bridge connecting relational and statistical learning.

Among the strong motivations for using a relational model is its ability to model dependencies between related instances. Intuitively, we would like to use our information about one object to help us reach conclusions about other, related objects. For example, in web data, we should be able to propagate information about the topic of a document to documents it has links to and documents that link to it. These, in turn, would propagate information to yet other documents. Many researchers have proposed a process along the lines of this relational "influence propagation" idea [3, 44, 32]. Chakrabarti et al. [3] describe a relaxation labeling algorithm that makes use of the neighboring link information. The algorithm begins with the labeling given by a text-based classifier constructed from the training set. It then uses the estimated class of neighboring documents to update the distribution of the document being classified. The intuitions underlying these procedural systems can be given declarative semantics using probabilistic graphical models [46, 15, 47].

1.4 Statistical Relational Learning

We refer to this emerging area of research as *statistical relational learning* (SRL). SRL research attempts to represent, reason, and learn in domains with complex relational and rich probabilistic structure. Other terms that have been used recently include probabilistic logic learning and multi-relational data mining. Many of the tasks known as structured prediction problems also overlap greatly with problems addressed by SRL research.

The majority of proposed SRL systems can be distinguished along several dimensions. The most common representation formalisms are based on either logic (e.g., rule-based formalisms) or frame-based (e.g., objected-oriented) formalisms.

The probabilistic semantics are mostly based on graphical models or stochastic grammars; early SRL approaches were often defined in terms of directed graphical models (e.g., Bayesian networks) whereas recently there has been a growing interest in undirected models (e.g., Markov networks). The directed models can represent complex generative models while the undirected models can represent non-causal dependencies. Other alternatives, such as dependency networks [19] and mixed directed and undirected models,6 are also possible.

The logical interpretation for most SRL languages (e.g., probabilistic relational models, Bayesian logic programs, relational Markov networks) is often in terms of least Hebrand models and the probabilistic semantics is most often in terms of a possible worlds semantics. Some of the early approaches, such as knowledge-based model construction (KBMC) [48], relied on procedural semantics. There are other possibilities, described in greater deal in the upcoming chapters.

The semantics of many of the SRL systems is given in terms of an unrolled or ground graphical model. Thus, one approach to doing inference in these models is to perform the appropriate probabilistic inference in the base-level model. One simple KBMC-style optimization is to make use of the query in the construction of the network. Rather than constructing the entire base-level model, the construction may be made more efficient by constructing only the portion of the network required to answer the query. But this doesn't exploit any of the inherent structure in the probabilistic model. Pfeffer et al. [38] observe that in many cases the models can be decomposed into loosely coupled systems, and show how the interfaces between the components can be used to encapsulate inference within the components. This allows the reuse and caching of inferences and can lead to significant improvements in efficiency during inference. More general approaches, such as first-order variable elimination [41, 1], combine variable elimination with unification and allow a lifted inference to be performed (see chapter 15 for details).

Not surprisingly, learning is a fundamental component in any SRL approach. The power of the structured representation is the hierarchical nature of the statistical models. The advantage of the hierarchical models, and what distinguishes them from "flat" statistical models, is parameter sharing or parameter tying. Parameter sharing occurs when potentially distinct parameters of the model are constrained to be the same. A simple example occurs in a hidden Markov model: because of the Markovian assumption, the parameters determining the next state are the same at each time instance, hence we do not require distinct parameters indexed by specific values of t, we simply have one set of parameters $\theta_{t+1|t}$.

This parameter tying not only gives us a compact model for rich classes of distributions but is also what enables robust parameter estimation to even be feasible. Unlike traditional ML scenarios, where the learning system is given as input a sequence of i.i.d. observations, the input to an SRL learning algorithm is most often just a single, richly connected, instance. If there were no parameter sharing, this instance would be of little use for performing statistical inference. But, because the same parameters are used in multiple places in the model, we can still extract meaningful statistics from the data to use in our statistical inference procedures.

Model selection is a challenging SRL problem. Similar to work in propositional graphical models, many approaches make use of some type of heuristic search through the model space. Methods for scoring propositional graphical models have been extended for SRL learning [12, 13]. The search can make use of certain biases defined over the model space, such as allowing dependencies only among attributes of related instances according to the entity relationship model or the use of binding patterns to constrain clauses to consider adding to the probabilistic rules.

Certain common issues arise repeatedly, in different guises, in a number of the SRL systems. One of the most common issues is feature construction and aggregation. The rich variety in structure combined with the need for a compact parameterization gives rise to the need to construct relational features or aggregates [12] which capture the local neighborhood of a random variable. Because it is infeasible to explicitly define factors over all potential neighborhoods, aggregates provide an intuitive way of describing the relational neighborhood. Common aggregates include taking the mean or mode of some neighboring attribute, taking the min or the max, or simply counting the number of neighbors. More complex, domain-specific aggregates are also possible. Aggregation has also been studied as a means for propositionalizing a relational classification problem [25, 23, 26] Within the SRL community, Perlich and Provost [36, 37] have studied aggregation extensively and Popescul and Ungar [42] have worked on statistical predicate invention.

Structural uncertainty is another common issue that researchers have begun investigating. Many of the early SRL approaches consider the case where there is a single logical interpretation, or relational skeleton, which defines the set of random variables, and there is a probability distribution over the instantiations of the random variables. Structural uncertainty supports uncertainty over the relational interpretation. Koller and Pfeffer [24] introduced several forms, including number uncertainty, where there is a distribution over the number of related objects. Getoor et al. [16] studied learning models with structural uncertainty, and showed how these representations could be supported by a probabilistic logic-based system [14]. Pasula and Russell [35] studied identity uncertainty, a form of structural uncertainty which allows modeling uncertainty about the identity of a reference. Most of these models rely on a closed world assumption to define the semantics for the models. More recently, Milch et al. [29] have investigated the use of nonparametric models which allow an infinite number of objects and support an open-world model (see the chapter 13 for details). Other recent flexible approaches include the infinite relational models of Kemp et al. [20] and Xu et al. [50].

1.5 Chapter Map

The book begins with several introductory chapters providing tutorials for the material which many of the later chapters build upon. chapter 2 is on graphical models and covers the basics of representation, inference, and learning in both directed and undirected models. Chapter 3 by Džeroski describes ILP. ILP, unlike many other ML approaches, has traditionally dealt with multi-relational data. The learned models are typically described by sets of relational rules called *logic programs*, and the methods can make use of logical background knowledge. Chapter 4 by Sutton and McCallum covers conditional random fields (CRFs), a very popular class of models for structured supervised learning. An advantage of CRFs is that the models are optimized for predictive performance on only the subset of variables of interest. The chapter provides a tutorial on training and inference in CRFs, with particular attention given to the important special case of linear CRFs. The chapter concludes with a discussion of applications to information extraction.

Then next set of chapters describes several frame-based SRL approaches. Chapter 5 provides an introduction to probabilistic relational models (PRMs). PRMs are directed graphical models which can capture dependencies among objects and uncertainty over the relational structure. In addition to describing the representation, the chapter describes algorithms for inference and learning. Chapter 6 describes Markov relational networks (RMNs), which are essentially CRFs lifted to the relational setting. A particularly relevant advantage of RMNs over PRMs is that acyclicity requirements do not hinder modeling complex, non-causal correlations concisely; however, as in the non-relational case, this comes at the price of more expensive parameter estimation. Another advantage of RMNs, like CRFs, is that they are well suited to discriminative training. Algorithms for inference and learning are given. Chapter 7, by Heckerman et al., describes a graphical language for probabilistic entity-relationship models (PERs). One of the contributions of this chapter is its discussion of the relationship between PERs, PRMs, and plate models. Plate models [2, 17] were introduced in the statistics community as a graphical representation for hierarchical models. They can represent the repeated, shared, or tied parameters in a hierarchical graphical model. PERs synthesize these approaches. The chapter describes a directed version of PERs, DAPERs, and gives a number of illustrative examples. Chapter 8, by Neville and Jensen, describes relational dependency networks (RDNs). RDNs extend propositional dependency networks to relational domains, and, like dependency networks, have some advantages over directed graphical models and undirected models. This chapter describes the representation, inference, and learning algorithms and presents results on several data sets.

The next four chapters describe logic-based formalisms for SRL. An introductory chapter, chapter 9 by Cussens, surveys this area, describing work on some of the early logic-based formalisms such as Poole's work on probabilistic Horn abduction [39] and independent choice logic [40], Ngo and Haddawy's work on probabilistic knowledge bases [34] and Sato's work on the PRISM system [43], and Ng and Subrahmanian's work on probabilistic logic programming [33]. Cussens compares and contrasts these approaches and describes some of the common representational issues, making connections to approaches described in later chapters. Chapter 10, by Kersting and De Raedt, describes Bayesian logic programs (BLPs). Their approach combines Bayesian networks and logic programs to "upgrade" them to a representation which overcomes the propositional nature of Bayesian networks and the purely logical nature of logic programs. This chapter gives an introduction to BLPs, describing both a Bayesian logic programming tool and a graphical representation for them. Chapter 11, by Muggleton and Pahlavi, describes stochastic logic programs (SLPs). SLPs were originally introduced as a means of extending the expressiveness of stochastic grammars to the level of logic programs. The chapter provides several example programs and describes both parameter estimation and structure learning. Chapter 12, by Domingos and Richardson, describes Markov logic. Markov logic combines Markov networks and first-order logic. First-order logic formulae are given weights; the formulae define a log-linear model with a feature for each grounding of the logical formulae with the appropriate weights. The relationship between many of the other SRL approaches and Markov logic networks (MLNs) is discussed, along with several common SRL tasks such as collective inference, link prediction, and object identification. Inference and learning in MLNs are presented.

Many of the approaches discussed so far have assumed, either implicitly or explicitly, several practical assumptions (the closed-world assumption, domain closure, unique names) about the underlying logical interpretation in order to define the underlying semantics. Chapter 13, by Milch et al. describes BLOG, a system especially tailored toward cases in which these assumptions are not appropriate. BLOG models define stochastic processes for generating worlds; inference in these models is done via a sampling process. Chapter 14, by Pfeffer, describes IBAL, a functional programming language for probabilistic AI. IBAL supports a rich decision-theoretic framework which includes probabilistic reasoning and utility maximization. The chapter describes the syntax and semantics for the IBAL, along with a sophisticated inference algorithm which exploits both lazy evaluation and memoization for efficient inference.

One of the issues that comes up in many of the approaches is the need to perform effective inference in large scale probabilistic models. Many of the approaches can make use of "lifted" inference, inference which is done at level of the first-order representation directly, rather than at the propositional level. Chapter 15 describes first-order variable elimination, an algorithm for lifted probabilistic inference, and presents recent results.

One of the issues that comes up in each of the learning algorithms is the need for feature generation and selection. Chapter 16, by Popescul and Ungar, examines this issue in the context of structured generalized linear regression (SGLR). They address the need for an integrated approach to feature generation and selection. Chapter 17, by Davis et al., addresses a related issue, the need for view learning to support feature generation and selection. They describe two approaches and present results on a mammography analysis system.

Chapter 18, by Fern et al. surveys recent work in reinforcement learning in relational domains. There has been a lot of recent work on relational learning within the reinforcement learning setting and our collection does not try to comprehensively cover its scope. Instead we have chosen a representative contribution describing a novel approach to approximate policy iteration which is applicable to very large relational Markov decision problems.

One of the domains which naturally lends itself to SRL techniques is natural language processing. Chapter 19 by Bunescu and Mooney shows how RMNs can be used for information extraction. An advantage of their approach is that inference and learning support "collective information extraction" in which dependencies between extractions are exploited. They present results on extracting protein names from biomedical abstracts. Chapter 20, by Roth and Yih, also investigates SRL approaches for information extraction, specifically for combining named entity and relation extraction. They show how a linear programming formulation can capture the required global inference.

1.6 Outlook

In this introduction we have touched on a number of the common themes and issues that will be developed in greater detail in the following chapters. While a single unified framework has yet to emerge, we believe that the book highlights the commonalities, and clarifies some of the important differences among proposed approaches. Along the way, important representational and algorithmic issues are identified.

Statistical relational learning is a young and exciting field. There are many opportunities to develop new methods and apply the tools to compelling real-world problems. We hope this book will provide an introduction to the field, and stimulate further research, development, and applications.

References

[1] R. Braz, E. Amir, and D. Roth. Lifted first-order probabilistic inference. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2005.

[2] W. Buntine. Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, 3:159–225, 1994.

[3] S. Chakrabarti, B. Dom, and P. Indyk. Enhanced hypertext categorization using hyperlinks. In *Proceedings of ACM International Conference on Management of Data*, 1998.

[4] D. Cohn and T. Hofmann. The missing link—a probabilistic model of document content and hypertext connectivity. In *Proceedings of Neural Information Processing Systems*, 2001.

[5] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery. Learning to extract symbolic knowledge from the World Wide Web. In *Proceedings of the National Conference on Artificial Intelligence*, 1998.

[6] J. Cussens. Loglinear models for first-order probabilistic reasoning. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 1999.

[7] L. Dehaspe, H. Toivonen, and R.D. King. Finding frequent substructures in chemical compounds. In *International Conference on Knowledge Discovery and Data Mining*, 1998.

[8] T. Dietterich and R. S. Michalski. Inductive learning of structural descriptions: Evaluation criteria and comparative review of selected methods. *Artificial Intelligence*, 16:257–294, 1986.

[9] S. Dzeroski and N. Lavrac, editors. *Relational Data Mining*. Kluwer, Berlin, 2001.

[10] P. Finn, S. Muggleton, D. Page, and A. Srinivasan. Discovery of pharmacophores using the inductive logic programming system Progol. *Machine Learning*, 30(1-2):241–270, 1998.

[11] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 2002.

[12] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1999.

[13] L. Getoor. *Learning Statistical Models from Relational Data*. PhD thesis, Stanford University, Stanford, CA, 2001.

[14] L. Getoor and J. Grant. PRL: A probabilistic relational language. *Machine Learning Journal*, 62(1-2):7–31, 2006.

[15] L. Getoor, E. Segal, B. Taskar, and D. Koller. Probabilistic models of text and link structure for hypertext classification. In *Proceedings of the IJCAI Workshop on Text Learning: Beyond Supervision*, 2001.

[16] L. Getoor, N. Friedman, D. Koller, and B. Taskar. Learning probabilistic models of link structure. *Journal of Machine Learning Research*, 3:679–707, 2002.

[17] W. Gilks, A. Thomas, and D. Spiegelhalter. A language and program for complex Bayesian modeling. *The Statistician*, 43:169–177, 1994.

[18] F. Hayes-Roth and J. McDermott. Knowledge acquisition from structural descriptions. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1997.

[19] D. Heckerman, D. Chickering, C. Meek, R. Rounthwaite, and C. Kadie. Dependency networks for inference, collaborative filtering and data visualization. *Journal of Machine Learning Research*, 1:49–75, 2000.

[20] C. Kemp, J. Tenenbaum, T. Griffiths, T. Yamada, and N. Ueda. Learning systems of concepts with an infinite relational model. In *Proceedings of the National Conference on Artificial Intelligence*, 2006.

[21] K. Kersting, L. De Raedt, and S. Kramer. Interpreting Bayesian logic programs. In *Proceedings of the AAAI-2000 Workshop on Learning Statistical Models from Relational Data*, 2000.

[22] R. King, S. Muggleton, A. Srinivasan, and M. Sternberg. Structure-activity relationships derived by machine learning: The use of atoms and their bond connectivities to predict mutagenicity by inductive logic programming. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 93, pages 438–442, 1996.

[23] A. Knobbe, M. deHaas, and A. Siebes. Propositionalisation and aggregates. In *Proceedings of the Fifth European Conference on Principles of Data Mining and Knowledge Discovery*, 2001.

[24] D. Koller and A. Pfeffer. Probabilistic frame-based systems. In *Proceedings of the National Conference on Artificial Intelligence*, 1998.

[25] S. Kramer, N. Lavrac, and P. Flach. Propositionalization approaches to relational data mining. In S. Dzeroski and N. Lavrac, editors, *Relational Data Mining*. Springer-Verlag, New York, 2001.

[26] M. Krogel, S. Rawles, F. Zeezny, P. Flach, N. Lavrac, and S. Wrobel. Comparative evaluation of approaches to propositionalization. In *Proceedings of the International Conference on Inductive Logic Programming*, 2003.

[27] N. Lavrac and S. Dzeroski. *Inductive Logic Programming: Techniques and Applications*. Ellis Horwood, New York, 1994.

[28] C. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, 1999.

[29] B. Milch, B. Marthi, S. Russell, D. Sontag, D. Ong, and A. Kolobov. BLOG: Probabilistic models with unknown objects. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2005.

[30] S. Muggleton, editor. *Inductive Logic Programming*. Academic Press, London, 1992.

[31] S. Muggleton. Learning stochastic logic programs. In *Proceedings of the AAAI-2000 Workshop on Learning Statistical Models from Relational Data*, 2000.

[32] J. Neville and D. Jensen. Iterative classification in relational data. In *Proceedings of the AAAI-2000 Workshop on Learning Statistical Models from Relational Data*, 2000.

[33] R. Ng and V.S. Subrahmanian. Probabilistic logic programming. *Information and Computation*, 101(2):150–201, 1992.

[34] L. Ngo and Peter Haddaway. Answering queries from context-sensitive probabilistic knowledge bases. *Theoretical Computer Science*, 171:147–171, 1997.

[35] H. Pasula and S. Russell. Approximate inference for first-order probabilistic languages. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2001.

[36] C. Perlich and F. Provost. Aggregation-based feature invention and relational concept classes. In *International Conference on Knowledge Discovery and Data Mining*, 2003.

[37] C. Perlich and F. Provost. Distribution-based aggregation for relational learning with identifier attributes. *Machine Learning Journal*, 62(1-2):65–105, 2006.

[38] A. Pfeffer, D. Koller, B. Milch, and K. Takusagawa. SPOOK: A system for probabilistic object-oriented knowledge representation. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 1999.

[39] D. Poole. Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence*, 64(1):81–129, 1993.

[40] D. Poole. The independent choice logic for modelling multiple agents under uncertainty. *Artificial Intelligence*, 94(1–2):5–56, 1997.

[41] D. Poole. First-order probabilistic inference. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2003.

[42] A. Popescul and L. Ungar. Structural logistic regression for link analysis. In *KDD Workshop on Multi-Relational Data Mining*, 2003.

[43] T. Sato. A statistical learning method for logic programs with distribution semantics. In *Proceedings of the International Conference on Inductive Logic Programming*, 1995.

[44] S. Slattery and T. Mitchell. Discovering test set regularities in relational domains. In *Proceedings of the International Conference on Machine Learning*, 2000.

[45] R. Stepp and R. S. Michalski. Conceptual Clustering: Inventing goal-oriented classifications of structured objects. Technical Report 940, University of Illinois, Urbana, 1985.

[46] B. Taskar, E. Segal, and D. Koller. Probabilistic classification and clustering in relational data. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2001.

[47] B. Taskar, P. Abbeel, and D. Koller. Discriminative probabilistic models for relational data. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2002.

[48] M. Wellman, J. Breese, and R. Goldman. From knowledge bases to decision models. *Knowledge Engineering Review*, 7(1):35–53, 1992.

[49] P. Winston. Learning structural descriptions from examples. In P. H. Winston, editor, *The Psychology of Computer Vision*, pages 157–209. McGraw-Hill, New York, 1975.

[50] Z. Xu, V. Tresp, K. Yu, and H. Kriegel. Infinite hidden relational models. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2006.