Artificial Intelligence 174 (2010) 147–151

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Book review

A. Darwiche, Modeling and Reasoning with Bayesian Networks, Cambridge, 2009.

1. Introduction

The field of probabilistic reasoning with Bayesian networks has experienced fruitful advancement in the last three decades, generating a large body of rich literature. Beginning with Pearl's influential 1988 book, there have been now over a dozen books that cover the diverse topics in this field. These books, successively published over a two-decade period, cover the latest development of the field at the time of the writing. Furthermore, as the field sufficiently matures to have a broader impact on AI practice, it is necessary to make the key results more accessible to a broader audience beyond researchers specialized in the field. As a result, newly published books increasingly appeal better to a wider readership. The book Modeling and Reasoning with Bayesian Networks by Adnan Darwiche is an excellent example of this trend. The book provides a comprehensive, accessible and advanced treatment of the representational and computational issues surrounding probabilistic reasoning with Bayesian networks. Adnan Darwiche is a leading researcher and is responsible for a number of techniques presented in the book.

The book contains eighteen chapters which can be partitioned, in my opinion, into six groups according to their topics, although Darwiche makes no such partition. Chapters in each group are not necessarily adjacent in the book.

- Fundamentals: 1, 2, 3, 4
- Modeling: 5, 16
- Exact inference: 6, 7, 8, 10, 12, 13
- Approximate inference: 14, 15
- Learning: 17, 18
- Theoretical issues: 9, 11

Below, we first summarize individual chapters and identify specific strengths and weaknesses. These summaries are organized according to the above groups. After that, we remark on the book as a whole.

2. Fundamentals

This group includes the first four chapters. Although the core materials contained are usually covered in other books on probabilistic reasoning, a difference in accessibility can be made. Hence, the comments below have paid more attention to the notational and presentational choices.

Chapter 1 introduces probabilistic reasoning and Bayesian networks informally through a brief overview of the history of automated reasoning. It starts with McCarthy's proposal for model-based systems and the early commitment to deductive logic. That is followed by non-monotonic logic to the rescue of limits of deduction and initial rejection of probabilistic approach. It highlights Pearl's historical contribution and proceeds to the new wave of probabilistic reasoning systems since the 1990s. Readers are well informed of why Bayesian networks are important to contemporary AI, not only historically but also on deep rational foundations.

No book dealing with probabilistic reasoning can completely escape the interpretation of probabilities (objective versus subjective). This issue is treated only lightly (less than half of page 6). No explicit stance is taken, while readers are reminded that the use of "degree of belief" in the book is not necessarily subjective. This caters better to diverse readers, but arguably appears lacking of coherence. Applications of probabilities are abundant in the book, some well-suited for the objective interpretation and others make sense only under the subjective interpretation. For coherence, subjective probabilities may be viewed as generalizing objective probabilities by assuming that upon learning the physical probability of an event, an agent would set its subjective probability accordingly (see [2], page 11).

The formal content commences at Chapter 2, which describes the propositional logic for representing and reasoning about events. Just enough materials, to be referred in later chapters, are presented clearly and concisely. For readers who have been exposed to propositional logic through the popular Russell and Norvig [5] text, note that some logic terms are

doi:10.1016/j.artint.2009.11.006

defined differently. A *model* there is called a *world* here, and the set of worlds that satisfy a sentence α is called the *models* of α. Entailment, denoted by $\models$, is defined as satisfaction between a world and a sentence. Yet, this seemingly weaker definition allows entailment between two sentences to be defined (page 17) semantically equivalently to a stronger notion of entailment; and the presentation is more concise and intuitive.

Chapter 3 describes probability calculus. The binary classification of worlds from propositional logic is extended by assigning degree of beliefs to them. The axioms of probability (referred to as properties of belief) are then derived through a series of intuitive arguments and examples. Belief updating is clearly presented through Bayes conditioning. Note that the term *faithful* is used with very different meaning than in the Bayesian network literature, e.g. [7], which corresponds to the notion of *perfect map* in Chapter 4. A comprehensive and insightful treatment on soft evidence is given at the end, which is the best presentation I have seen on the topic.

Bayesian networks are introduced in Chapter 4, followed by graphoid axioms for probabilistic independence. The notion of d-separation is presented as the criterion for graphical separation in directed acyclic graphs. The chapter is concluded with relations between probabilistic independence and graphical separation in terms of I-map, D-map and P-map.

A novel convention is adopted in the definition of the numerical component of a Bayesian network. Instead of denoting the CPT associated with each variable X with parents U by probability $\Pr(X|U)$ as usual, the notation $\Theta_{X|U}$ is used to emphasize that it is part of the parametrization. Such convention arguably improves clarity and accessibility for a number of concepts. The presentation of the chain rule, for example, now has the joint probability distribution on the left-hand side, and the product of parameters on the right-hand side: a familiar pattern where a target entity is defined in terms of model parameters. This improved clarity extends to sensitivity analysis (Chapter 16), where the impact of changes in parameters on the values of posterior probabilities is analyzed (rather than the impact of changes in some probabilities on the values of other probabilities), and to learning (Chapters 17 and 18), where these parameters are learned from data (and then used to derive probabilities).

The $I(X, Z, Y)$ notation for conditional independence, and the similar notation for graphical separation, were originally adopted in Pearl's 1988 book. Pearl switched to an alternative notation in his later book [4]. I am glad to see that the original notations are used by Darwiche as they are both symbolic (hence concise) as well as graphical (hence intuitive).

3. Modeling

This group includes Chapters 5 and 16 and they are particularly important and useful to knowledge engineers.

Chapter 5 focuses on construction of Bayesian networks and approaches the task from the knowledge representation perspective (as opposed to the machine learning perspective). Four types of queries are defined, which are frequently encountered in probabilistic reasoning using Bayesian networks: probability of evidence (PR), prior and posterior marginals (MAR),[1] most probable explanation (MPE), and maximum a posteriori hypothesis (MAP). Issues in network construction in relation to these queries are then discussed and illustrated through a number of example applications: medical diagnosis, trouble-shooting (including intermittent faults), system reliability, channel coding (including convolutional codes modeled with dynamic Bayesian networks), common sense reasoning, and genetic linkage analysis (with an accessible background explanation).

Chapter 16 deals with sensitivity analysis and considers the impact of parameters in Bayesian networks on the results of probabilistic queries. Two types of sensitivity analysis are considered: query robustness and query control. Query robustness is concerned with possible changes in the results of queries given the changes to network parameters. Both network-independent robustness and network-specific robustness are treated. Query control addresses the inverse problem and identifies parameter changes that are necessary and sufficient to produce a particular change in the result of a query.

4. Exact inference

This group contains Chapters 6, 7, 8, 10, 12, 13 and is, in my opinion, the central part of the book.

Chapters 6, 7 and 8 present the three main classes of algorithms for exact inference in Bayesian networks, with the focus on PR and MAR queries. Chapter 6 presents a variable elimination algorithm for MAR queries and its variation, bucket elimination. The pruning of a Bayesian network based on the query structure is also described.

The jointree algorithm for MAR queries is covered in Chapter 7. This is through a much accessible derivation which starts by extending variable elimination into factor/potential elimination, continues with a series of intuitive intermediate cluster trees for message passing, and results in the jointree algorithm. Historically, major versions of the jointree algorithm were discovered between 1988 and 1990, while major versions of the variable elimination method were discovered after 1993. Hence, the derivation of the joint tree algorithm through variable elimination is both interesting and innovative.

Two major variations of the jointree algorithm, the Shenoy–Shafer and the Hugin architectures are described at the end of Chapter 7. The presentation of the Shenoy–Shafer architecture as rooted message passing (page 167) is, however, inaccurate. In fact, under the typical Shenoy–Shafer architecture, each node "... is allowed to send its message to a particular neighbor as soon as it has messages from all its other neighbors" [6] (page 45). Hence, no root needs to be selected and there are no

[1] Abbreviations PR and MAR are adopted by this reviewer.

inward and outward phases defined based on the root selected. Such a message passing mechanism is *asynchronous* [9] and more flexible (less global control) in comparison with the *rooted* mechanism used in the typical Hugin architecture.

Chapter 8 discusses exact inference by conditioning. After cutset conditioning, recursive conditioning is covered in depth. Two extreme versions (relative to caching) of recursive conditioning for PR queries are introduced. The no-caching extreme has the best space complexity but the worst time complexity, and the full-caching extreme drops the time complexity but increases the space complexity. An any-space version follows that allows controlled tradeoff of time with space. The graphical structure, called a *dtree*, used for PR queries is then extended into a *dgraph* for MAR queries.

Chapter 10 continues with exact inference methods with focus on MPE and MAP queries. Two classes of algorithms are described based on variable elimination and systematic search, respectively.

In Chapter 12, Bayesian networks are converted into arithmetic circuits for answering PR queries, which are further extended to answer MPE queries. The contributions covered are mostly attributed to Darwiche and collaborators. They have not been described in other books on probabilistic reasoning, and the chapter makes these useful techniques more accessible. The conversion is termed "compilation" and it is emphasized that the process takes place offline and only once. Such emphasis seems unjustified when it is only applied to the conversion into arithmetic circuits. For instance, couldn't the conversion of Bayesian networks into jointrees or dtrees take place offline and only once as well? The formulation of arithmetic circuits bears much similarity to query DAGs [1] proposed by Darwiche and Provan in late 1990s. It is unclear why the relation between query DAGs and arithmetic circuits is not discussed in bibliographic remarks.

Inference methods based on variable elimination, jointree, and conditioning are *structure-based* as their complexity is dependent only on the graphical structure of the Bayesian network, not the values of numerical parameters. Chapter 13 explores structures among parameters to allow for more efficient inference than what is supported by structure-based algorithms. This class of inference techniques has not be considered systematically in other books on probabilistic reasoning. It includes techniques to encode Bayesian network into CNFs and to refine conditioning and elimination algorithms.

5. Approximate inference

Approximate inference is covered in Chapters 14 and 15 with the objective of providing the basis to manage the tradeoff between inference accuracy and efficiency.

Inference methods presented in Chapter 14 are based on belief propagation (message passing) and are divided into two classes. One class is based on loopy propagation, where belief propagation that works exactly in singly connected Bayesian networks are iteratively applied to multiply connected networks. Appendix C offers some background on iterative methods and will help readers to gain insight into these methods. The other class is based on edge-deletion, where some edges in a Bayesian network are deleted so that exact inference methods can be efficiently applied to the resultant network.

Readers would be better informed if some results, currently missing, are included which are, in my opinion, quite relevant to loopy propagation. One of them is Murphy et al. [3] which reports the first empirical study on loopy propagation in a diverse set of Bayesian networks. This study is significant for two reasons. Prior work (as quoted in the book) was based mainly on Bayesian networks for coding applications, while networks used in [3] have more diverse structures and parametrizations. Furthermore, the result from prior studies is primarily positive, while the result reported by Murphy et al. is mixed. Other relevant work [8] (Section 3.5) classifies cycles in graphical models and shows that message passing in *non-degenerate cycles* is fundamentally limited. Loopy propagation has been actively studied for about a decade and Chapter 14 can be viewed as a state-of-art coverage of the progress. As acknowledged by Darwiche (at multiple points in the chapter), there has been no general method that guarantees the convergence of loopy propagation and a bounded error when the propagation does converge. The study in [8] sheds light to why this is so.

Chapter 15 presents approximate inference methods based on stochastic sampling, including Rao–Blackwell sampling, logic sampling, likelihood weighting, particle filtering, and Gibbs sampling. Aspects of estimation theory are introduced at the start and are used to unify the presentation and analysis of the approximate methods. This is unique among comparable books and provides additional insight into these algorithms.

6. Learning

Chapters 17 and 18 are dedicated to learning Bayesian networks from data. They compliment methodologies for constructing Bayesian networks from the knowledge representation perspective (Chapter 5) by tackling the task from the machine learning perspective. Algorithms for learning Bayesian networks are classified into three general approaches: the maximum likelihood approach, the Bayesian approach, and the constraint-based approach. The first two are covered, while the third is not treated in the book.

The maximum likelihood approach is described in Chapter 17, which covers parameter learning as well as structure learning from both complete and incomplete data. The maximum likelihood classification provides a novel perspective to unify a number of well-known methodologies such as the EM algorithm, Chow and Liu's tree algorithm, MDL scoring, greedy and optimal structure search.

The Bayesian approach to learning Bayesian networks is covered in Chapter 18. Meta-networks for learning Bayesian networks are introduced as a representation to encode prior knowledge over parameters as well as structures, and to incorporate data. The task of learning a Bayesian network from a given dataset is then cast as inference in a corresponding meta-network. Methods for learning parameters given structure and for learning both structure and parameters are treated. Overall, this chapter is not as accessible in comparison with the rest: the materials are more condensed and there are insufficient examples and illustrations.

7. Theoretical issues

This group contains Chapters 9 and 11. Although they both include algorithmic components, they primarily deal with fundamental issues of theoretical importance.

Chapter 9 deeply examines the relationship amongst inference methods presented in Chapters 6, 7 and 8. From the perspective of graph decomposition, these algorithms decompose the Bayesian network according to elimination order, jointree, and dtree, respectively. It is shown that the construction of an optimal elimination order, an optimal jointree, or an optimal dtree are all equivalent to the optimal triangulation of a graph. This treatment provides an insightful unification of these seemingly different inference methods.

The content of Chapter 11 can be split into two halves. The first half analyzes the complexity of decision problems (D-PR, D-MAR, D-MPE, and D-MAP) that correspond to the four types of probabilistic queries (PR, MAR, MPE, and MAP). The other half presents reductions of PR and MPE queries to weighted model counting on CNF sentences and weighted MAXSAT, respectively. So far, books on probabilistic reasoning have not included any comprehensive analysis of the above decision problems. Darwiche carries out the analysis in a comprehensible manner which is easy to follow. On the other hand, the introduction and motivation to the analysis could be improved. On a minor note, the description on correspondence between the reasoning problems and decision problems (page 270) could be improved for better clarity. More importantly, it is not explained why readers who are interested in how to answer PR, MAR, MPE, and MAP queries and the complexity of doing so should care about the complexity of D-PR, D-MAR, D-MPE, and D-MAP.

8. Other remarks

Given the richness of the literature on probabilistic reasoning with Bayesian networks, it is impossible to treat all relevant topics in depth within a single volume. Chapter 1 acknowledges a number of topics that are either uncovered in the book or only covered to a limited degree, and points readers to further references. These topics include the philosophical foundation of probability theory, decision theory, Bayesian networks with continuous variables, undirected graphical models, causality, and relational and first-order modeling. One closely related topic can be added to the uncovered list—multiagent Bayesian networks. The treatment of Bayesian networks in the book essentially assumes (implicitly) a single-agent paradigm. That is, a single computational agent is equipped with a model of its environment in the form of a Bayesian network, either inserted into the agent by human knowledge engineers or acquired through learning. The agent makes observations over its environment, updates its belief on the environment by reasoning with the Bayesian network, and answers queries, either self-induced or from the human principal. The single-agent paradigm is sufficient for many applications, but there are other applications where it is undesirable or difficult or impossible to assemble the necessary knowledge into a single Bayesian network and to grant its access to any single agent for inference. A multiagent paradigm for Bayesian networks is better suited for these applications. For a multiagent treatment of Bayesian networks, see [8].

In writing this book, Darwiche was "... driven ... by a strong desire to provide the most intuitive explanations" (Preface) to the representational and computational techniques treated in the book. He certainly succeeded in achieving this goal. The book is both practical and advanced. The first five chapters are sufficient for students and practitioners to gain the necessary knowledge in order to build Bayesian networks for moderately sized applications with the aid of a software tool. Such software tools, such as Samlam (Chapter 5), are readily available commercially or freely from many research labs around world. All major inference methods are covered in later chapters which allow researchers and software developers to implement their own software systems tailored to their needs. All important algorithms are presented in a uniform pseudocode format, which facilitates understanding and implementation. Most concepts and methods are motivated and introduced intuitively before defined and specified formally, with their computational properties rigorously analyzed. It is a comprehensive book that can be used for self study by students and newcomers to the field or as a companion for courses on probabilistic reasoning. Experienced researchers may also find deeper information on some topics. In my opinion, the book should definitely be in the bookshelf of everyone who teaches Bayesian networks and builds probabilistic reasoning agents.

References

[1] A. Darwiche, G. Provan, Query DAGs: A practical paradigm for implementing belief-network inference, in: E. Horvitz, F. Jensen (Eds.), Proc. 12th Conf. on Uncertainty in Artificial Intelligence, Portland, OR, 1996, pp. 203–210.

[2] K. Korb, A. Nicholson, Bayesian Artificial Intelligence, Chapman and Hall, 2004.

[3] K.P. Murphy, Y. Weiss, M.I. Jordan, Loopy belief propagation for approximate inference: An empirical study, in: K.B. Laskey, H. Prade (Eds.), Proc. 15th Conf. on Uncertainty in Artificial Intelligence, Stockholm, 1999, pp. 467–475.

[4] J. Pearl, Causality: Models, Reasoning, and Inference, Cambridge University Press, 2000.

[5] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, Prentice Hall, 2003.

[6] G. Shafer, Probabilistic Expert Systems, Society for Industrial and Applied Mathematics, Philadelphia, 1996.

[7] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, 2nd ed., Springer Lecture Notes in Statist., MIT Press, Cambridge, MA, 2000.
[8] Y. Xiang, Probabilistic Reasoning in Multiagent Systems: A Graphical Models Approach, Cambridge University Press, Cambridge, UK, 2002.
[9] Y. Xiang, F.V. Jensen, Inference in multiply sectioned Bayesian networks with extended Shafer–Shenoy and lazy propagation, in: Proc. 15th Conf. on Uncertainty in Artificial Intelligence, Stockholm, 1999, pp. 680–687.

Yang Xiang
Department of Computing and Information Science,
University of Guelph, Canada
E-mail address: yxiang@cis.uoguelph.ca

6 July 2009
Available online 17 November 2009