International Journal of Approximate Reasoning 51 (2010) 800–819

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Arc refractor methods for adaptive importance sampling on large Bayesian networks under evidential reasoning

Haohai Yu *, Robert van Engelen [1]

Computer Science Department, Florida State University, Tallahassee, FL 32306, USA

ARTICLE INFO

Article history:
Received 30 July 2009
Received in revised form 19 March 2010
Accepted 27 April 2010
Available online 5 May 2010

Keywords:
Uncertainty
Probabilistic reasoning
Bayesian networks
Approximate Bayesian inference

ABSTRACT

Approximate Bayesian inference by importance sampling derives probabilistic statements from a Bayesian network, an essential part of evidential reasoning with the network and an important aspect of many Bayesian methods. A critical problem in importance sampling on Bayesian networks is the selection of a good importance function to sample a network's prior and posterior probability distribution. The initially optimal importance functions eventually start deviating from the optimal function when sampling a network's posterior distribution given evidence, even when adaptive methods are used that adjust an importance function to the evidence by learning. In this article we propose a new family of Refractor Importance Sampling (RIS) algorithms for adaptive importance sampling under evidential reasoning. RIS applies "arc refractors" to a Bayesian network by adding new arcs and refining the conditional probability tables. The goal of RIS is to optimize the importance function for the posterior distribution and reduce the error variance of sampling. Our experimental results show a significant improvement of RIS over state-of-the-art adaptive importance sampling algorithms.

1. Introduction

The *Bayesian Network* (BN) [1] formalism is one of the dominant representations for modeling knowledge and uncertainty in intelligent systems [2,3]. Popular applications can be found in areas of expert systems, decision support systems, bioinformatics, medicine, image processing, and information retrieval. A BN is a probabilistic graphical model of a *joint probability distribution* over a set of *statistical variables* and their probabilistic independence relationships. By utilizing these relationships, the *posterior probability distribution* can be efficiently calculated for *evidential reasoning* to answer probabilistic queries about the variables and their influences.

Evidential reasoning by exact probabilistic inference is NP-hard [4]. Exact inference for *belief updating* [1] is only efficient for networks of limited complexity [5] and is not feasible for large or complex models. Approximations are also NP-hard [6]. However, approximate methods have so-called *anytime* [7] and *anywhere* [8] properties that make them preferable over exact methods for realistic large-scale BN applications.

Stochastic simulation algorithms, *stochastic sampling* or *Monte Carlo* (MC) methods, form one of the most prominent classes of approximate inference algorithms [3]. *Logic sampling* [9] was the first and simplest sampling algorithm proposed. *Likelihood weighting* [10] was designed to overcome the poor performance of logic sampling under evidential reasoning with unlikely evidence. *Markov Chain Monte Carlo* (MCMC) [11,12] methods form another important group of stochastic sampling

* Corresponding author.
E-mail addresses: hy06@fsu.edu (H. Yu), engelen@cs.fsu.edu (R. van Engelen).
[1] Member of IEEE, ACM Senior Member.

0888-613X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2010.04.006

algorithms, e.g. *Gibbs sampling*, *Metropolis sampling*, and *hybrid-MC sampling* [13–15]. *Stratified sampling* [16], *hypercube sampling* [17], and *quasi-MC* [18] methods generate random samples from uniform distributions to improve sampling results.

Approximate Bayesian inference algorithms based on *importance sampling* [19] are among the most frequently used. Examples are *Self Importance Sampling* (SIS) and *Heuristic Importance Sampling* [20]. Importance sampling algorithms differ in the choice of *importance function* that defines the sampling distribution. The closer the importance function is to the actual distribution (with the possible exception of heuristics to support thicker tails [21]), the better the performance of sampling to yield good convergence results in acceptable time [12]. A well-known problem is that the importance functions, which are all initially based on the prior distribution, eventually start diverging from the *exact posterior distribution* under evidential reasoning. The error variance of importance sampling may significantly increase when evidence variables are instantiated and the distribution's prior (in) dependence relations change to posterior (in) dependence relations.

To address this problem, one approach is to approximate the posterior conditional probability distribution by *mini-bucket elimination* [22], see for example [23,24]. *Dynamic Importance Sampling* (DIS) [25,26] extends this approach by using *probability trees* [27], that leverages the context-specific independence relations, and by refining the approximated *conditional probability tables* (CPTs) of a Bayesian network during sampling.

Other advances in importance sampling are based on *learning methods* to learn an importance function under evidential reasoning, e.g. *adaptive IS* [28] and *Adaptive Importance Sampling* (AIS-BN) [29], or methods that directly compute an importance function based on the prior distribution and the evidence, e.g. *Evidence Pre-propagation Importance Sampling* (EPIS-BN) [30,31]. For learning, both SIS and AIS-BN adapt the importance function by making quantitative adjustments to the CPTs of the BN using a particular *CPT learning* algorithm. However, it has been shown [32] that CPT learning does not suffice to match the adjusted importance function to the posterior distribution, because the divergence error of all CPT learning-based importance functions are bounded from below.

More recently, restructuring methods for adaptive importance sampling that change both the network structure and CPTs have been proposed. In [31] a *factorization algorithm* is described[2] that adjusts the BN structure to further reduce the error variance of importance sampling under evidential reasoning. Also DIS [25,26] restructures the *potential functions* through the variable elimination process. However, a drawback of network restructuring is that potentially many new arcs can be introduced. This is especially problematic for large networks, because the complexity of the network topology may be significantly increased and this can lead to large CPTs. The number of table entries for a vertex is exponential in the in-degree of the vertex.

In this article we present a new family of *arc refractor methods* for adaptive importance sampling algorithms that prevent an arc explosion after network restructuring, while ensuring that the importance function still closely approximates the posterior distribution. We show that even a small number of arc refractors can lead to a significant improvement of importance sampling under evidential reasoning. Our technique, *Refractor Importance Sampling* (RIS), locally modifies the network structure by way of "refracting" arcs incident to evidence vertices away to a subset of the ancestors of the evidence vertices. RIS attempts to closely approximate the importance sampling function of the posterior distribution and, as a consequence, the error variance of RIS is not bounded from below. Our preliminary investigation has shown that RIS yields better convergence results in most cases [32].

An additional benefit of RIS is that it can effectively reduce the negative effects of *sample rejection*, a well-known problem of importance sampling on BN with non-strictly positive distributions [29,33]. To approach this problem, we use a RIS-based network modification to propagate zero probabilities in CPTs upwards to the roots of the network to filter invalid samples more efficiently.

The remainder of this article is organized as follows. Section 2 introduces the theoretical basis for importance sampling on BN with RIS. The family of restructuring methods for adaptive importance sampling based on RIS arc refractors is presented in Section 3. Section 4 augments the RIS approach by limiting sample rejection when sampling non-strictly positive BN. An empirical validation of the approach is given in Section 5, using synthetic and real-world BNs. We compare RIS with other related work in Section 6. Finally, Section 7 summarizes our results and conclusions.

2. Preliminaries

We start with the theoretical basis of the BN formalism and importance sampling on BN, see also [1,3,19,21,33].

2.1. Bayesian networks

We use uppercase letters for variables and lowercase letters for the states of the variables. Boldface letters are used for sets of variables (vertices), arcs, or states. The relation $X \to Y$ represents a directed arc between two vertices X and Y in a graph G. The reflexive, transitive closure $\xrightarrow{*}$ of the arc relation $\to$ represents directed paths between vertices in G.

Definition 1. A *Bayesian network* $BN = (G, \Pr)$ consists of a directed acyclic graph (DAG) $G = (\mathbf{V}, \mathbf{A})$ with vertices $\mathbf{V}$, arcs $\mathbf{A} \subseteq \mathbf{V} \times \mathbf{V}$, and a joint probability distribution Pr over the discrete random variables $\mathbf{V}$ (represented by the vertices of G). Pr is defined by

[2] We will refer to Algorithm 1 in [31] as the "factorization algorithm."