Anytime Lifted Belief Propagation

Rodrigo de Salvo Braz ∗   BRAZ@AI.SRI.COM
Sriraam Natarajan †   NATARASR@BIOSTAT.WISC.EDU
Hung Bui ∗   BUI@AI.SRI.COM
Jude Shavlik †   SHAVLIK@CS.WISC.EDU
Stuart Russell ‡   RUSSELL@CS.BERKELEY.EDU

∗ SRI International, Menlo Park, California, USA

† Department of Computer Science, University of Wisconsin, Madison, USA

‡ Computer Science Division, University of California, Berkeley, USA

Abstract

Lifted first-order probabilistic inference, which manipulates first-order representations directly, has been receiving increasing attention. To date, all lifted inference methods require a model to be *shattered* against itself and evidence, before inference starts. In many situations this produces a new model that is not far from propositionalized, therefore canceling the benefits of lifted inference. We present an algorithm, Anytime Lifted Belief Propagation, that corresponds to this intuition by performing shattering *during* belief propagation inference, on an as-needed basis, starting on the most relevant parts of a model first. The trade-off is having an (exact) bound (an interval) on the query's belief rather than an exact belief. Bounds are useful when approximate answers are sufficient and, in decision-making applications, can even be enough to determine the decision that would be picked from the exact belief. Moreover, the bounds can be made to converge to the exact solution as inference and shattering converge to the entire model. Interestingly, this algorithm mirrors theorem-proving, helping to close the gap between probabilistic and logic inference.

1. Introduction

First-order probabilistic models are specified with potential or conditional probability templates parameterized by logical variables. Following Poole (2003), we call these templates *parfactors*, for *parameterized factors*. Two examples of parfactors are

$$\phi(p(X), q(X, Y))$$
$$P(advises(Prof, St)|ta\_for(St, Prof)), St \neq john,$$

where ϕ and P stand for potential and conditional probability functions that apply to each instantiation of the parameterized random variables (*atoms*) by the quantified *typed logical variables* X, Y, St and $Prof$. Logical variables often have constraints on them (here, with equality formulas only). The semantics of a set of parfactors is the graphical model formed by their instantiations satisfying their constraints.

Lifted inference on first-order probabilistic models (that is, inference that manipulates and keeps the first-order structure, avoiding extensive propositionalization) has been receiving increasing attention recently (Poole, 2003; de Salvo Braz et al., 2007; Milch et al., 2008; Singla & Domingos, 2008). To date, all lifted inference methods require a model to be *shattered* against itself and evidence, before inference starts. Shattering means dividing the random variables of the model into clusters of exactly symmetric variables, that is, variables with the same probabilistic statements on them and their neighbors and therefore exhibiting identical behavior. Evidence is often provided at the level of random variables on specific individuals, typically causing all random variables involving them to form singleton clusters. For many problems this is very close to propositionalization, and

the gains from lifted inference are greatly decreased.

The reason shattering is needed in advance is because the algorithms that have been lifted (belief propagation and variable elimination) do require the entire model in order to compute a query's belief. So in general the entire model needs to be used, requiring it to be entirely shattered. However recent work on *box propagation* (Mooij & Kappen, 2008) shows how to derive *bounds* on beliefs from using only a portion of a model. This allows us to *gradually* shatter the model while obtaining useful bounds on the query. Interestingly, this also corresponds to the intuition that reasoning only consider sub- or individual cases in an as-needed basis, as it is done in theorem proving where unification and resolution are gradually used.

We present an algorithm that corresponds to this intuition by performing lifted belief propagation (Singla & Domingos, 2008), but shattering the model *during* belief propagation inference, starting on the most relevant parts of a model first. The method uses box propagation to provide an (exact) bound, that is, an interval guaranteed to contain the query's belief rather than an exact belief. Bounds are useful when only an approximate answer is needed, or when beliefs are used for supporting decisions. In the latter, they may suffice for determining the decision that would be picked from the exact beliefs. Moreover, the bounds can be made to converge to the exact belief as inference and shattering proceed to include the entire model.

Interestingly, this algorithm mirrors logic theorem proving. In fact, when parfactors are hard constraints, it *reduces* to theorem proving. Moreover, the closer they are to hard constraints, the closer the behavior of the algorithm is to theorem proving. This has several advantages. It shows that the algorithm is viable even if much of a model is purely logic; helps closing the gap between probabilistic and logic inference; and produces higher-level, more intelligible reasoning traces akin to proof trees.

2. The Algorithm

Our algorithm is a combination of box propagation (Mooij & Kappen, 2008), and lifted belief propagation (BP) (Singla & Domingos, 2008).

Box propagation (shown in Fig. 1 on a factor network) works by considering only a subset S of the model that is increasingly expanded from the query outwards (evidence is represented by factors with potential zero on assignments inconsistent with it). At every step from (b) to (d), factors are included in the set so as to complete some random variable's blanket (we do not show
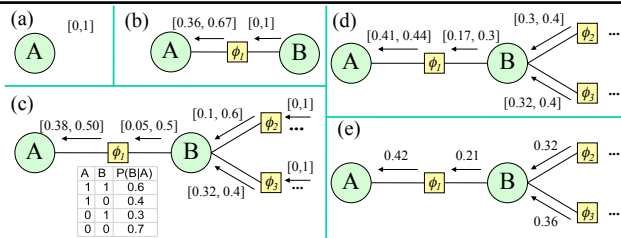

Figure 1. Box propagation on a binary variables network.

the expansions from (d) to (e), however, only their consequences). We include the table for factor ϕ_1 but omit the tables for other factors. For simplicity, this paper uses binary variables only, but all algorithms in it can work with multi-valued variables.

When a blanket of a variable V is completed, the messages coming to new factors from outside S are bounded by $[0, 1]$, and a bound on V is calculated by considering these input message bounds. The resulting bound is then propagated all the way to the query, narrowing the bound on its belief. Thus this is an anytime algorithm since a bound on the query is always available and becoming narrower as processing continues. A bound is computed by simply considering the extremes of input bounds and recording the extreme values of the output of the potential function (for factor nodes) or their product (for variable nodes). After enough expansions (either by including all factors or by using enough of them), we converge to 0-width bounds with exact messages, as in (e). Note that the algorithm works for loopy belief propagation by creating new nodes for variables and factors when they are found more than once, creating an *unrolled* network.

Lifted belief propagation is based on the idea that symmetric variables (that is, variables with exactly the same set of dependencies) will receive and generate the same belief messages. It determines these sets (called *supernodes*) by shattering as a pre-processing step, and performs message passing between them.

Anytime Lifted BP works by using only a subset of the model for box propagation, but with supernodes, as in Lifted BP. Shattering is performed only as it is needed for accommodating the parfactors brought it at each step, thus minimizing it. Figure 2 shows a detailed example. We do not write the algorithm in detail due to space restrictions.

Box propagation carries with it the need for updating a message bound out of a node when one of its input message bounds is updated. This makes its application to first-order probabilistic models even more significant, since they are typically specified with highly regular potential functions based on logical formulas or
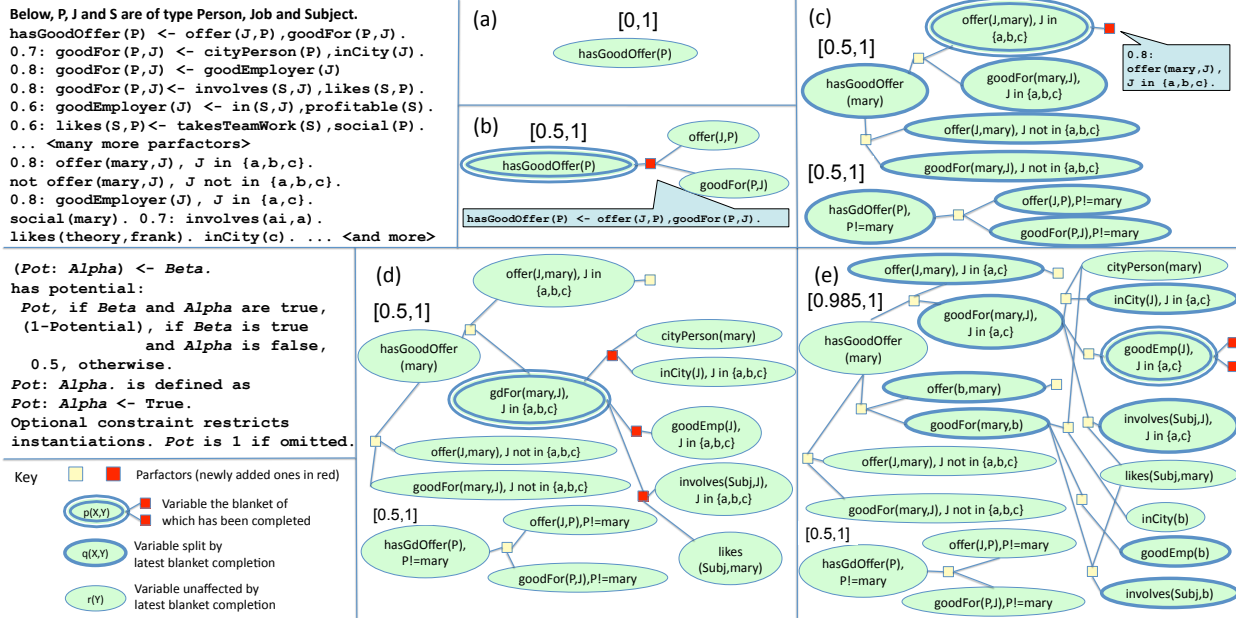
Figure 2. An example of Anytime Lifted BP in action. Note how all jobs, then later a, b and c, and still later only a and c, are grouped together at several stages, even though their random variables are not symmetric in the complete model. This model would be completely grounded by regular lifted inference. Once we reach the parfactors indicating that both a and c are likely to have made offers to Mary and to be good employers, the bound on her having a good offer becomes narrow, without examining many other facts that also bear influence on it. Doing so would close the gap. Note that the parfactor specification notation is irrelevant to the algorithm, which applies to other representations such as MLNs (Singla & Domingos, 2008).

aggregation functions, which have linear, rather than exponential, update costs (we omit details).

3. Conclusion

We presented an algorithm that performs lifted first-order probabilistic inference without shattering the entire model in advance. This allows for faster useful inference when bounds are enough (such as for decision-making). Importantly, it allows us to treat classes of random variables as a group even when they are not exactly symmetric; in other words, it allows us to reason in general about objects that are only *approximately* symmetric, where the notion of approximation becomes more restricted the more precise an answer is required. When objects in the model are hypotheses in a learning problem, then the decision-making scenario naturally brings up a type of active learning where the most relevant data is sought first. An interesting extension in that direction would be making the algorithm consider the cost of obtaining evidence or sharper bounds on certain variables. The algorithm also mirrors theorem proving, helping to close the gap between logic and probabilistic reasoning.

References

de Salvo Braz, R., Amir, E., & Roth, D. (2007). Lifted first-order probabilistic inference. In L. Getoor and B. Taskar (Eds.), *An introduction to statistical relational learning*, 433–451. MIT Press.

Milch, B., Zettlemoyer, L., Kersting, K., Haimes, M., & Kaelbling, L. P. (2008). Lifted probabilistic inference with counting formulas. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI-2008)*. Chicago, Illinois, USA.

Mooij, J. M., & Kappen, H. J. (2008). Bounds on marginal probability distributions. *Advances in Neural Information Processing Systems 21 (NIPS*2008)* (pp. 1105–1112).

Poole, D. (2003). First-order probabilistic inference. *Proceedings of the 18th International Joint Conference on Artificial Intelligence* (pp. 985–991).

Singla, P., & Domingos, P. (2008). Lifted first-order belief propagation. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI-2008)*. Chicago, Illinois, USA.