Hierarchical Marginalization

Anonymous Author(s)
Affiliation
Address
email

Abstract

The exactness of hierarchical partitioning is fairly easy to intuit, but proving it is mildly more difficult. Using the 'port node' intuition given in the main paper, the method to construct the proof becomes visible. We marginalize over a set of port nodes, and each sub-graph is marginalized over it's port nodes plus the ones it shares in common, and so on.

The existence of a message passing algorithm for this task is somewhat trivial, since if we have a cutset it'll split the graph into trees, for which there is an exact algorithm. When implementing such an algorithm, just make sure you don't marginalize over the same tree twice.

A Exactness of sub-graph partitioned marginalization

A.1 Definitions

Definition 1 (Factor graph)**.** A *factor graph* $G = (F, X)$ is a bipartite graph representing a factorization of a joint probability distribution over a set of variables X. It consists of:

- A set of variable nodes $X = \{x_1, x_2, \ldots, x_n\}$ with domains $\mathcal{X}_{x_i}$.
- A set of factor nodes $F = \{\phi_1, \phi_2, ...\phi_m\}$ where $\phi_i : \Omega(\Theta(\phi_i)) \to \mathbb{R}^+$.
- A function $\Theta : F \to \mathbb{P}(X)$[1] that maps from a factor to the subset of variables it's dependent on.

Additionally, providing extra arguments to a factor doesn't change it's output:

$$\phi(x) = \phi(y) : x \in \Omega(X), y \in \Omega(X \cup Y)$$

Definition 2 (Instantiation)**.** An instantiation is a mapping from a set of variables X to one variable in each one's domain. The function Ω returns a set of all possible instantiations given a set of variables.

$$\Omega(Y) = \{\omega : (\forall y \in Y, \omega(y) \in \mathcal{X}_y), (\forall S \subseteq Y : \omega[S] = \{\omega(s) : s \in S\}\}$$

An instantiation can map a variable to it's assigned value $\omega(y)$, or map a set of variables to their assigned values $\omega(S)$

Definition 3 (Factorized joint distribution)**.** Given a factor graph G, the joint distribution over X is given by $P_G : \Omega(X) \to [0, 1]$ where:

[1] where $\mathbb{P}(X)$ is the power set of X

$$\Phi_G(x) = \prod_{\phi \in F} \phi(x)$$
$$Z = \sum_{x \in \Omega(X)} \Phi_G(x)$$
$$P_G(x) = \Phi_G(x)/Z$$

The behavior when provided extra arguments is the same for a single factor:

Proof. For $y \in \Omega(X \cup Y), x \in \Omega(X)$

$$\begin{aligned} \Phi_G(y) &= \prod_{\phi \in F} \phi(y) \\ &= \prod_{\phi \in F} \phi(x) \\ &= \Phi_G(x) \end{aligned}$$

$$\begin{aligned} P_G(y) &= \Phi_G(y)/Z \\ &= \Phi_G(x)/Z \\ &= P_G(x) \end{aligned}$$

Definition 4 (Marginal distribution)**.** Given a subset of variables $A \subseteq X$ the *marginal distribution* of A is defined as:

$$\psi_{G,A}(a) = \sum_{x \in \Omega(X \setminus A)} \Phi_G(x \cup a)$$
$$Z = \sum_{a \in \Omega(A)} \psi_{G,A}(a)$$
$$p_{G,A}(a) = \psi(a)/Z$$

where $a \in \Omega(A)$

A.2 Sub-graph factorization

A.2.1 Sub-graphs and near-disjoint-ness

Definition 5 (Sub-graph of a factor graph)**.** A *sub-graph* $\bar{G} = (\bar{X}, \bar{F})$ of a factor graph $G = (X, F)$ is a factor graph formed by a subset of variables $\bar{X} \subseteq X$ and a subset of factors $\bar{F} \subseteq F$, such that every factor in $\bar{F}$ is defined only on variables from $\bar{X}$.

Definition 6 (Near-disjoint sub-graphs)**.** A collection of sub-graphs $\{\bar{G}_i = (\bar{X}_i, \bar{F}_i)\}_i$ of G is *near-disjoint* w.r.t. A if:

$$\bigcup_i \bar{F}_i = F$$
$$\forall_{i \neq j}, \bar{F}_i \cap \bar{F}_j = \emptyset$$
$$\forall_{i \neq j}, \bar{X}_i \cap \bar{X}_j \subseteq A$$

The factor sets partition F. The sets of variables $\bar{X}_i$ may share variables in A, but are otherwise disjoint

A.2.2 Sub-graph factorization

Lemma 1 (Sub-Graph Joint Distribution Factorization)**.** Suppose $G = (X, F)$ is a factor graph and $\{\bar{G}_i = (\bar{X}_i, \bar{F}_i)\}_i$ is a collection of near-disjoint sub-graphs partitioning F. Then the joint distribution can be factorized:

$$\begin{aligned}
\Phi_G(x) &= \prod_{\phi \in F} \phi(x) \\
&= \prod_i \prod_{\phi \in \bar{F}_i} \phi(x) && \text{(assoc. \& comm. mul.)} \\
&= \prod_i \Phi_{\bar{G}_i}(x) && \text{(Def. 3)}
\end{aligned}$$

$$\begin{aligned}
Z &= \sum_{x \in X} \Phi_G(x) \\
P_G(x) &= \Phi_G(x)/Z
\end{aligned}$$

A.2.3 Marginalization with near-disjoint sub-graphs

Lemma 2 (Generalized Distributive Law [1])**.** Let $(\mathcal{K}, \sum, \prod)$ be a commutative semiring. Let $\{D_i\}_{i=1}^k$ be pairwise–disjoint finite sets and let $f_i : D_i \to \mathcal{K}$. Then

$$\sum_{(x_1,\ldots,x_k) \in \prod_{i=1}^k D_i} \prod_{i=1}^k f_i(x_i) \;=\; \prod_{i=1}^k \sum_{x_i \in D_i} f_i(x_i)$$

"sum over the Cartesian product of products" equals "product of the individual sums"

Theorem 1 (Marginal factorization)**.** Given a subset $A \subseteq X$ and a factor graph $G = (X, F)$ that partitions into near-disjoint sub-graphs $\{\bar{G}_i = (\bar{X}_i, \bar{F}_i)\}_i$ that share variables only in A, the marginal distribution of A can be factorized as:

$$\begin{aligned}
\psi_{G,A}(a) &= \sum_{x \in \Omega(X \setminus A)} \Phi_G(x \cup a) && \text{Def. 4} \\
&= \sum_{x \in \Omega(X \setminus A)} \prod_i \Phi_{\bar{G}_i}(x \cup a) && \text{Lem. 1} \\
&= \sum_{x \in \Omega(X \setminus A)} \prod_i \Phi_{\bar{G}_i}(x[\bar{X}_i] \cup a) && \text{Def. 2, 3, 6} \\
&= \prod_i \sum_{x_i \in \Omega(X_i \setminus A)} \Phi_{\bar{G}_i}(x_i \cup a) && \text{Lem. 2} \\
&= \prod_i \psi_{\bar{G}_i,A}(a) && \text{Def. 4}
\end{aligned}$$

$$\begin{aligned}
Z &= \sum_{a \in A} \psi_{G,A}(a) \\
p_{G,A}(a) &= \psi_{G,A}(a)/Z
\end{aligned}$$

The key insight is that because the sub-graphs $\bar{G}_i$ are near-disjoint except for A, the joint summation over $X \setminus A$ factorizes into independent summations over each $\bar{X}_i \setminus A$. Each of these summations yields a marginal distribution $\psi_{\bar{G}_i,A}(a)$. Combining these pieces leads to the factorization of the marginal distribution into a product of sub-graph marginals.

A.2.4 Hierarchical partitioning marginalization

When a suitable partition of G is not immediately available (i.e., the graph remains connected even after removing A), introduce a new set of variables P such that for which $P \cup A$ there is a partitioning of near-disjoint sub-graphs.

Definition 7 (Partition)**.** This partition C is a set of variables that, when conditioned upon (i.e., fixed), partitions the factor graph into a collection of near-disjoint sub-graphs. Formally, if $Q \subseteq X$, then:

$$\psi_{G,A}(a) = \sum_{c \in \Omega(C)} \psi_{G,A \cup C}(a \cup c).$$

The summation of a partition involves marginalizing w.r.t A^C over a marginalized distribution over $(A \cup C)^C$. Partitions ensure that each sub-graph induced by $Q \cup C$ is near-disjoint, allowing factorization techniques to apply.

Theorem 2 (Hierarchical factorization via partitions)**.** Consider a factor graph G and a chosen partition C. If conditioning on $Q \cup C$ makes the sub-graphs near-disjoint, then:

$$\psi_{G,Q}(q) = \sum_{c \in \Omega(C)} \psi_{G,Q \cup C}(q \cup c).$$

For a given partition, the sub-graphs can themselves be partitioned into near-disjoint sub-sub-graphs with independent partitioners. The partitioners unique to each sub-graph therefor can be summed out independently for each sub-graph.

$$\begin{aligned}
\psi_{G,A}(a) &= \prod_i \psi_{\bar{G}_i,A}(a) \\
&= \prod_i \psi_{\bar{G}_i,A \cup \bar{C}_i}(a) \\
&= \prod_i \sum_{c \in \bar{C}_i} \psi_{\bar{G}_i,A \cup \bar{C}_i}(a \cup c).
\end{aligned}$$

The summation $\sum_c$ is where "summing out a sub-graph" comes from

Proof. Applying the previous factorization theorems to each sub-graph yields the factorization into products of marginals. Introducing additional nested partitions $\bar{C}_i$ for each sub-graph repeats the argument at a finer level of granularity, leading to hierarchical factorization similar to *Anytime Exact Belief Propagation*[2]. □

B Existence of exact message passing sub-graph marginalization algorithm

Definition 8 (Cycle Cutset Conditioning[3])**.** For any graph there exists a cutset on a factor graph such that belief propagation with an outer summation on the cutset is exact.

$$\psi_{G,A}(a) = CCC_{G,A}(a) = \sum_{c \in \Omega(C)} BP_{G,C \cup A}(c \cup a)$$

Lemma 3 (Nested Factor Lemma)**.** Marginalizing over a single factor defined as $= \psi_{\bar{G}_i,A}$ is equivalent to $\psi_{\bar{G}_i,A}$

By 1 as so long as the sub-graph marginalizations are exact, there exists an exact super-graph marginalization.

$$\psi_{G,A}(a) = \prod_i \bar{\phi}_{\bar{G}_i,A}$$

It can be shown that marginalizing over a single factor defined as $= \psi_{\bar{G}_i,A}$ is equivalent to $\psi_{\bar{G}_i,A}$.

$$\phi_{\bar{G}_i,A}(a) = \psi_{\bar{G}_i,A}$$

$$\begin{aligned}
\bar{\psi}_{\bar{G}_i,A}(a) &= \phi_{\bar{G}_i,A}(a) \\
&= \psi_{\bar{G}_i,A}(a)
\end{aligned}$$

Theorem 3 (Hierarchical Cutset Conditioning)**.** By 3 there exists an equivalent factor-graph for any sub-graph partitioning. And by 8 there exists a message passing algorithm for any marginalization over a factor-graph. Therefor, there exists an exact message passing algorithm between sub-graphs.

Proof by induction.

1. base case is regular cycle cutset conditioning (Def. 8)
2. any sub-graph partitioning can be turned into a sub-graph factor-graph (Lem. 3)
3. the cycle cutset conditioning algorithm can then be applied to the sub-graph factor graph and to the sub-graph factors.

References

[1] S. Aji and R. McEliece. The generalized distributive law. *IEEE Transactions on Information Theory*, 46(2):325–343, 2000.

[2] G. A. Ferreira, Q. Bertrand, C. Maussion, and R. de Salvo Braz. Anytime exact belief propagation. *CoRR*, abs/1707.08704, 2017.

[3] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.