HOP-MAP: Efficient Message Passing with High Order Potentials

Daniel Tarlow, Inmar Givoni, Richard Zemel
University of Toronto, Dept. of Computer Science
Motivation

• Modeling interactions among variables in undirected graphical models over discrete variables
• Pairwise models are convenient to use

But the pairwise models are not good enough
  – Even when optimum can be found, results are often not satisfying
Goal

• We need better models and convenient tools
  – Capture high order interactions
• In the general case – intractable, but useful classes exist where messages can be computed in poly-time.
• Setup - factor graphs over binary variables
  – Easy 1-of-N transformation from discrete variables

![Diagram of high order factor graph and max product belief propagation](image)
In this work...

• Develop new class of *order-based* potentials.

• Improve computation for existing class of *cardinality* potentials.

• Show how to easily create *compositions* of HOPs.

• **Code** available
Cardinality Potentials

\[ \theta(h) = f\left( \sum_{h_j \in h} h_j \right) \]

Function value based on number of on variables

<table>
<thead>
<tr>
<th>Related Work for General Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gupta et al. (2007)</td>
</tr>
<tr>
<td>Potetz et al. (2008)</td>
</tr>
<tr>
<td>Tarlow et al. (2008)</td>
</tr>
<tr>
<td>This work</td>
</tr>
</tbody>
</table>
Specialized Cardinality Potentials

• Pattern-based potentials: generalized Potts model (Kohli et al. 07).
• b-of-N potential (Huang & Jebara 07).
• Distributional priors, e.g., Dirichlet process, Pitman-Yor (Tarlow et al. 08).
Order-Based Potentials 1 – Convex Set

\[ f(h_1, \ldots, h_N) = \begin{cases} 
0 & \text{if } h_i = 1 \land h_k = 1 \Rightarrow h_j = 1 \forall i < j < k \\
-\alpha & \text{otherwise} 
\end{cases} \]
Order Based Potentials 2- Before-After

\[
f(x_1, \ldots, x_N, y_1, \ldots, y_N) = \begin{cases} 
0 & \text{if } \min_i x_i = 1 \text{ and } \max_j y_j = 1 \\
-\alpha & \text{otherwise}
\end{cases}
\]
Composition of Potentials

- If some variables act as logical switches
  \[ f(h_1, \ldots, h_N) = \begin{cases} 
  g_0(h_2, \ldots, h_N) & \text{if } h_1 = 0 \\
  g_1(h_2, \ldots, h_N) & \text{if } h_1 = 1
  \end{cases} \]
- And if \( g_1, g_2 \) are tractable HOPs
- Can efficiently compute messages based on known messages for tractable HOPs
- No need to re-derive updates
- Generalizes to >1 subset of switching variables \( h_s \) switching among \( K \) HOPs of (say) at most \( \log N \) cost with 
  \[ O(2^{|h_s|} + K \log N) \]
"Switch" variables $h_s$

Gates Representation

(Minka & Winn 08)

Auxiliary Variable $g(h_s)$

Composite Factor

"Switch" variables
What are these HOPs good for?

• Many existing and novel models can be constructed by mixing and matching of HOPs.
• Interactive Poster! Can you figure out which?
• Priors for image segmentation tasks:
  – Over segments size (multilabel)
  – Encourage % of pixels to be on (bg/fg)
  – Convex image parts
• Bipartite matching
• Affinity Propagation/Facility location
• User preferences of document $i$ over $j$ in ranking
• ...
Related Work

• **Context Specific Independence** (Boutilier et al. 96)
  – Original motivation: more structure than just graph
  – HOP-MAP generalizes tree-based CPT representation

• **COMPOSE** (Duchi et al. 07)
  – Same big computational idea: special purpose algorithms for computing max-product messages
  – HOP-MAP is finer-grained ➔ broadly applicable.
  – Can be used together (code for COMPOSE pending)

• **Gates** (Minka & Winn 08)
  – Composite factors essentially implement max-product for Gates over high order potentials.
Experiments – Image segmentation

- unary, pairwise
- unary, pairwise, convexity along
- unary, pairwise, \( f(k) = -\frac{1}{4}d^2 - k \)
- unary, pairwise, cardinality
- unary, pairwise, cardinality, convexity
Experiments – Rank Aggregation

• Document ranking for IR
• Goal – learn score for document-query pairs, then rank using scores.
• users are better at saying “I prefer doc \( i \) over doc \( j \)” than “I give it a score of 8.5” with respect to some query.
• Can be incorporated using before-after potentials.
• Simulated data
  – Unary potentials – high score -> high rank
  – \( \lambda \) : before-after potential strength
  – \( \rho \): fraction of satisfied constraints
Experiments – Rank Aggregation

\[ \lambda = 0.01 \]
\[ \rho = \frac{13}{18} \]

\[ \lambda = 0.05 \]
\[ \rho = \frac{15}{18} \]

\[ \lambda = 0.25 \]
\[ \rho = \frac{18}{18} \]
Conclusions

• HOPs of interest in real-world problems are often structured, and therefore are often tractable
• We are building a vocabulary to expand beyond pairwise interactions, conveniently.
• Present efficient/novel components for “inner loop” of message passing
• Can use “outer loop” of choice
  – Max-Product, GEMPLP (Globerson and Jaakkola, 2008)
    High-Order TRW due to Komodakis and Paragois, 2009)
• Code and Demos at http://www.cs.toronto.edu/~dtarlow/hops/