Exact and Approximate Inference in Associative Hierarchical Networks using Graph Cuts

Chris Russell1, Ľubor Ladický1, Pushmeet Kohli2, Philip H.S. Torr1

1 Oxford Brookes University
2 Microsoft Research, Cambridge

Associative Hierarchical Networks and related models

Formulation and inference

Experimental Evaluation on MSRC dataset

Standard CRF (Lafferty ‘01)
- Lacks long range interactions
- Local features are less informative
- Results often over smoothed

Cluster-based CRF (eg. Batra et al. CVPR08)
- Make an a priori decision which variables take the same label.
- Informative features based on aggregate properties of cluster
- Enforces consistency at a single larger scale
- Fixes the quantisation, may fail to separate entities or over separate.
- No way to recover from a misleading clustering

Higher order Robust PN model (Kohli, et al. – CVPR08)
- Potentials enforce weak consistency over clusters
- Multiple clusterings
- Robust to misleading clusters
- Limited to unary and pairwise potentials over variables
- PN potentials take the form of a robust majority voting

Associative Markov Network (Taskar et al ‘04)
- Potentials enforce weak consistency over clusters
- Multiple clusterings
- Robust to misleading clusters
- Limited to unary and pairwise potentials over variables
- PN potentials take the form of a robust majority voting

Associative Hierarchical Networks (AHNs) (Ladicky et al. – ICCV09)
- Encourages consistency at all scales
- Supports potentials defined over many differently sized regions
- Simultaneously resolves problem at all scales, minimizing inconsistency
- Robust to misleading segmentation
- Robust PN potentials reformulated as pairwise hierarchical energy
- New auxiliary variables take a value from the label-set or label free
- Unary and pairwise potentials at every quantization
- Multiple hierarchies of potentials
- Limited to PN-like associative interlayer connections

Unary Potential
\[ E^{(0)}(x) = \sum_{i} \psi_{0}^{(i)}(x^{(i)}) + \sum_{i,j} \psi_{0}^{(i,j)}(x^{(i)}, x^{(j)}) + \min_{x^{(i)}} E^{(0)}(x^{(i)}, x^{(j)}) \]

Pairwise Potential
\[ E^{(1)}(x^{(i)}, x^{(j)}) = \sum_{i} \sum_{j} \psi_{1}^{(i,j)}(x^{(i)}, x^{(j)}) + \min_{x^{(i)}} E^{(1)}(x^{(i)}, x^{(j)}) \]

Higher order Potential
\[ E^{(n)}(x^{(i)}, x^{(j)}) = \sum_{i} \sum_{j} \psi_{n}^{(i,j)}(x^{(i)}, x^{(j)}) + \min_{x^{(i)}} E^{(n)}(x^{(i)}, x^{(j)}) \]

Graph Cuts
- Standard CRF
- Cluster-based CRF
- Tree Model
- Deep Belief Net
- Hierarchical Network

The expressibility of common models versus AHNs

PN Model

Cluster-based CRF

Tree Model

Deep Belief Net

Hierarchical Network

Experimental Evaluation on MSRC dataset

Inference of the base layer is NP hard
\[ E(x) = \sum_{i} \psi_{0}(x^{(i)}) + \sum_{i,j} \psi_{0}(x^{(i)}, x^{(j)}) \]

but good graph-cut based approximations exist (Boykov et al. ’01)

Approximating higher order functions is much harder. Existing approximation schemes provide an O(n) approximation (n is the size of the largest clique) or are unbounded for the robust PN and AHN (Gould et al. ’09).

We remove the min terms from the equation to derive a pairwise energy,
\[ E(x, x^{(-)}) = \sum_{i} \psi_{0}(x^{(i)}) + \sum_{i,j} \psi_{0}(x^{(i)}, x^{(j)}) + E^{(1)}(x^{0}, x^{(-)}) \]

We reparameterise this energy function which allows us to perform bounded inference on the resulting energy, as on the base layer.

We then project this solution back into the original space, preserving the bounds.

This transformation from higher order to pairwise energy introduced new local minima which ‘trap’ move making optimisations.

We propose a new class of move-making algorithms which guarantee that all local minima found must be a local minima of the original higher order energy.