On Partial Optimality in Multi-label MRFs

P. Kohli[1] A. Shekhovtsov[2] C. Rother[1]
V. Kolmogorov[3] P. Torr[4]

[1]Microsoft Research Cambridge

[2]Czech Technical University in Prague

[3]University College London

[4]Oxford Brookes University

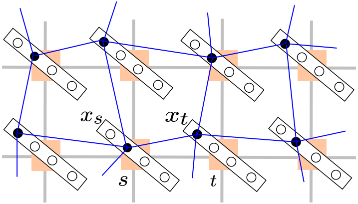
ICML, 2008

Outline

- Energy minimization $\boxed{\min_{\mathrm{x}} E(\mathrm{x}|\theta)}$ (MAP inference in MRF/CRF)

$$E(\mathrm{x}|\theta) = \sum_s \theta_s(x_s) + \sum_{st} \theta_{st}(x_s, x_t)$$

variables $x_s \in \mathcal{L} = \{1 \dots K\}$



- NP-hard in general
- Consider:

  conventional linear relaxation

  relaxation of a binarized problem

- Goal: study relations

Linear Programming Relaxation Approach

Relaxation LP-1

$$E(x|\theta) = \sum_s \theta_s(x_s) + \sum_{st} \theta_{st}(x_s, x_t) = \langle \theta, \mu(x) \rangle,$$

$$[\mu(x)]_s(i) = \delta_{\{x_s=i\}} \qquad [\mu(x)]_{st}(i,j) = \delta_{\{x_s=i\}}\delta_{\{x_t=j\}}$$

$$\min_{x \in \mathcal{L}^\mathcal{V}} \langle \theta, \mu(x) \rangle = \min_{\substack{A\mu=b \\ \mu\in\{0,1\}^n}} \langle \theta, \mu \rangle \geq \min_{\substack{A\mu=b \\ \mu\in[0,1]^n}} \langle \theta, \mu \rangle$$

- proposed many times independently [Schlesinger-76, Koster-98, Chekuri-00, Wainwright-03, Cooper-07]
- large-scale LP problem
- sub-optimal dual solvers [Koval-76, Wainwright-03, Kolmogorov-05]
- subgradient dual solvers [Schlesinger & Giginyak- 07, Komodakis *et al.*-07]

Binary Problems

- $\mathcal{L} = \{0, 1\}$ – pseudo-Boolean optimization [Boros, Hammer, ...]
- still NP-hard
- LP-relaxation (roof-dual) can be solved via network flow
- Can identify assignments which are persistent for all (some) optimal solutions

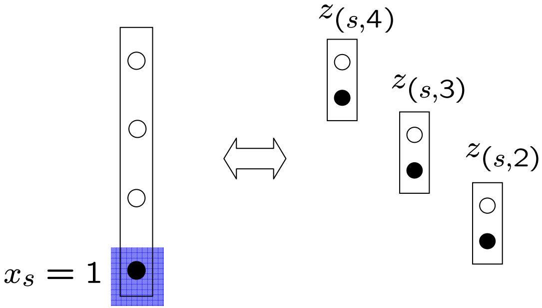
Definition

Relation (*e.g.* $x_s = \alpha$) is strongly persistent
if it is satisfied for all minimizers x.

Reduction to Binary Problem

$$E(\mathrm{x}|\theta) = \sum_s \theta_s(x_s) + \sum_{st} \theta_{st}(x_s, x_t)$$

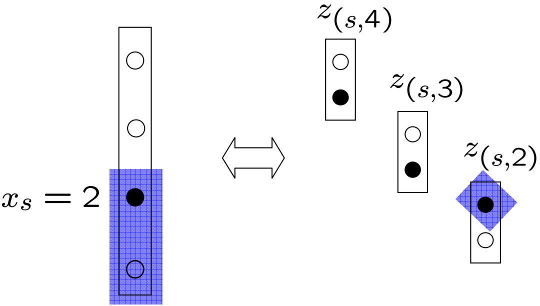
- Introduce $z_{(s,i)} = \delta_{\{i \leq x_s\}}$
  [Ishikawa-03, Kovtun-04, Schlesinger & Flach-06]



Reduction to Binary Problem

$$E(\mathrm{x}|\theta) = \sum_s \theta_s(x_s) + \sum_{st} \theta_{st}(x_s, x_t)$$

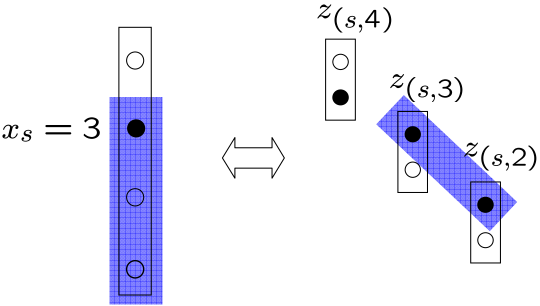
- Introduce $z_{(s,i)} = \delta_{\{i \le x_s\}}$
  [Ishikawa-03, Kovtun-04, Schlesinger & Flach-06]



Reduction to Binary Problem

$$E(\mathrm{x}|\theta) = \sum_s \theta_s(x_s) + \sum_{st} \theta_{st}(x_s, x_t)$$

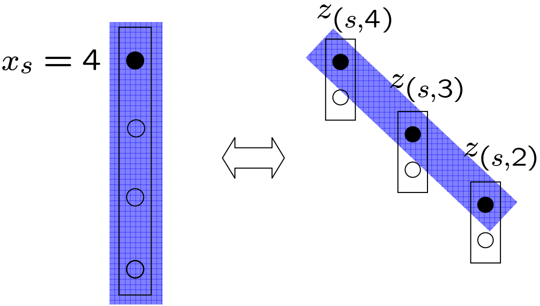
- Introduce $z_{(s,i)} = \delta_{\{i \leq x_s\}}$
  [Ishikawa-03, Kovtun-04, Schlesinger & Flach-06]



Reduction to Binary Problem

$$E(\mathrm{x}|\theta) = \sum_s \theta_s(x_s) + \sum_{st} \theta_{st}(x_s, x_t)$$

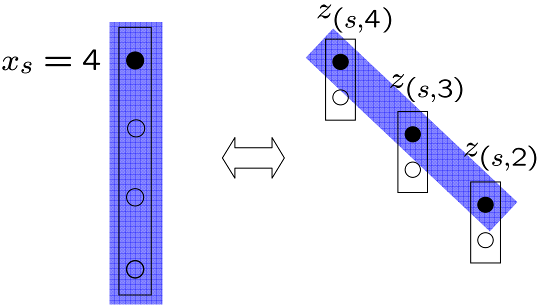
- Introduce $z_{(s,i)} = \delta_{\{i \leq x_s\}}$
[Ishikawa-03, Kovtun-04, Schlesinger & Flach-06]



Reduction to Binary Problem

$$E(\mathrm{x}|\theta) = \sum_s \theta_s(x_s) + \sum_{st} \theta_{st}(x_s, x_t)$$

- Introduce $z_{(s,i)} = \delta_{\{i \le x_s\}}$
  [Ishikawa-03, Kovtun-04, Schlesinger & Flach-06]

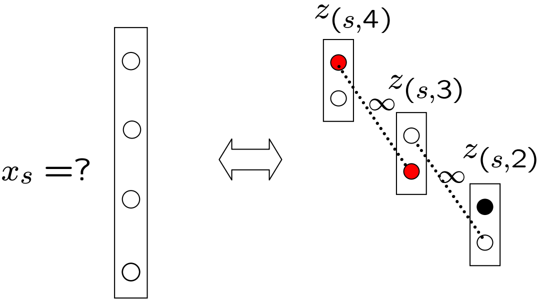


$$E(\mathrm{x}|\theta) = E(\mathrm{z}|\eta) = H(\mathrm{z}) + \sum_u \eta_u z_u + \sum_{uv} \eta_{uv} z_u z_v + \eta_{\text{const}}$$

Reduction to Binary Problem

$$E(\mathrm{x}|\theta) = \sum_s \theta_s(x_s) + \sum_{st} \theta_{st}(x_s, x_t)$$

- Introduce $z_{(s,i)} = \delta_{\{i \leq x_s\}}$
  [Ishikawa-03, Kovtun-04, Schlesinger & Flach-06]



$$E(\mathrm{x}|\theta) = E(\mathrm{z}|\eta) = H(\mathrm{z}) + \sum_{u \in V} \eta_u z_u + \sum_{uv \in A} \eta_{uv} z_u z_v + \eta_{\mathrm{const}}$$

Reduction to Binary Problem

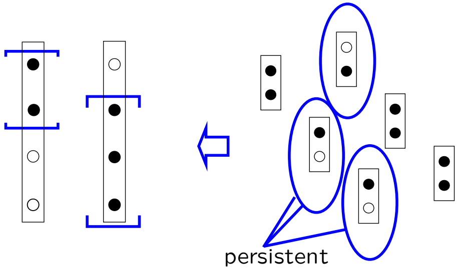
$$E(\mathrm{x}|\theta) = \sum_s \theta_s(x_s) + \sum_{st} \theta_{st}(x_s, x_t)$$

- Introduce $z_{(s,i)} = \delta_{\{i \leq x_s\}}$
  [Ishikawa-03, Kovtun-04, Schlesinger & Flach-06]

Relaxation LP-2 (roof-dual)

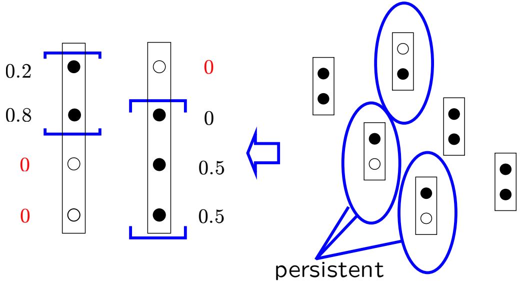
- Apply conventional LP-relaxation to the **binarized** problem $E(z|\eta)$
- Yields relaxation of the **original** problem

Persistencies in Multi-Label



- Hard constraints imply that non-persistent labels form intervals
- problem restriction / part of optimal solution

Persistencies in LP-1



Theorem

We show that persistency derived from LP-2 holds for LP-1 relaxation

Submodular Problems

Definition

Function $f : \mathcal{L}^{\mathcal{V}} \to \mathbb{R}$ is called submodular if
$f(\mathrm{x} \vee \mathrm{y}) + f(\mathrm{x} \wedge \mathrm{y}) \leq f(\mathrm{x}) + f(\mathrm{y}) \quad \forall \mathrm{x}, \mathrm{y} \in \mathcal{L}^{\mathcal{V}}$

- $(\mathrm{x} \vee \mathrm{y})_s = \max(x_s, y_s)$
- $(\mathrm{x} \wedge \mathrm{y})_s = \min(x_s, y_s)$

Subclass on which LP-2 = LP-1

Consider $E(\mathrm{x}|\theta) = \sum_s \theta_s(x_s) + \sum_{st} \theta_{st}(x_s, x_t)$

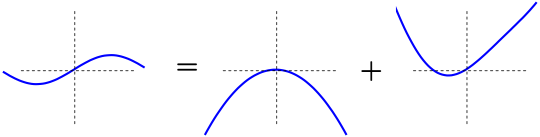
Theorem

If each $\theta_{st}(\cdot,\cdot)$ is submodular or supermodular, then LP-2 = LP-1

- LP-1 for this subclass can be solved using network flow model
- we have not found applications.

Submodular+Supermodular

- Decompose $E(\mathrm{x}|\theta) = E(\mathrm{x}|\theta^{\mathrm{sub}}) + E(\mathrm{x}|\theta^{\mathrm{sup}})$



- $\min_x E(x|\theta) \geq \min_x E(\mathrm{x}|\theta^{sub}) + \min_x E(x|\theta^{\mathrm{sup}})$ – (computable LB for bipartite graphs)

Statement

Tightest bound $=$ LP-2

- *c.f.* [Wainwright *et al.*-03] decomposition with trees.

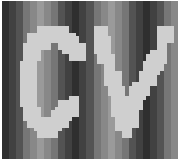
Experiments

Methods:

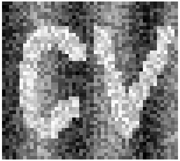
- derive restriction intervals $[x_s^{\min},\ x_s^{\max}]$ on the problem variables using network flow model for $LP-2$ (MQPBO)
- some variables get determined exactly – use
- apply other methods on restricted problem (MQPBO+X)
- derive more persistent constraints by probing (MQPBO-P)

Experiments

For some instances global minimum can be found



Original



Noisy Image

MQPBO-P (E=65382)

BP (E=65424)

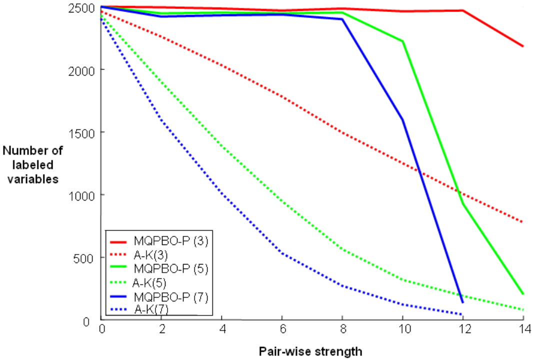
TRW-S (E=65398)

Expansion (E=65386)

Experiments

Random instance: how many variables are determined exactly?

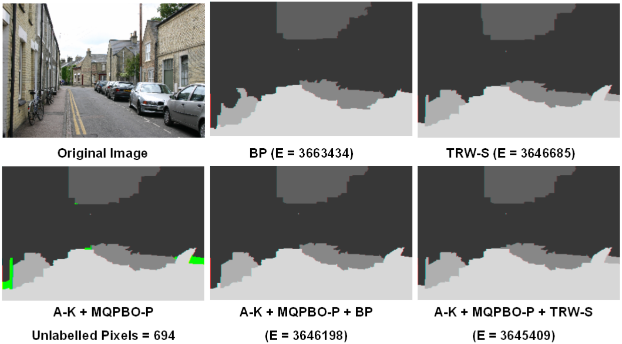
50×50 variables, comparison with [Kovtun-03]



Experiments

Real Instance: combined methods

Object segmentation and recognition model [Shottonet *al*.-05]



Conclusion

- There could be different low-order linear relaxations
- We studied some relations between two of them
- Dependence on the Ordering

  We assumed $\mathcal{L} = \{1, \ldots, K\}$ – ordered
  Order of labels for each variable x_s can be selected differently – exponentially many
  Order-independent reductions are possible, we investigated one and it has degenerate LP-relaxation solutions

Boros, E., & Hammer, P. (2002).
Pseudo-boolean optimization.
Discrete Applied Mathematics, 155–225.

Werner, T. (2007).
A linear programming approach to max-sum problem: A review.
PAMI, *29*, 1165–1179.

Schlesinger, D., & Flach, B. (2006).
Transforming an arbitrary minsum problem into a binary one (Technical Report TUD-FI06-01).
Dresden University of Technology.

Wainwright, M., Jaakkola, T., & Willsky, A. (2003).
Exact MAP estimates by (hyper)tree agreement.
In *Advances in neural information processing systems 15*, 809–816.

Boros, E., Hammer, P. L., & Tavares, G. (2006).

Preprocessing of unconstrained quadratic binary optimization
(Technical Report RRR 10-2006).
RUTCOR.