http://linqs.cs.umd.edu

Hinge-loss Markov Random Fields: Convex Inference for Structured Prediction

Stephen H. Bach, Bert Huang, Ben London, and Lise Getoor
University of Maryland, College Park

Introduction

- Hinge-loss Markov random fields are **powerful** models for **structured prediction**
- New **scalable** MPE inference algorithm much faster than inference in discrete MRFs
- **State-of-the-art performance** on four diverse learning tasks

Hinge-loss Markov Random Fields

Undirected probabilistic graphical models analogous to discrete MRFs

- Variables are **continuous valued** in [0,1]
- Potentials are **hinge-loss** functions
- Arbitrary linear **constraints**

$$P(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z} \exp\left[-\sum_{j=1}^{m} \lambda_j \max\left\{\ell_j(\mathbf{Y},\mathbf{X}), 0\right\}^{p_j} \right]$$

where $\ell_j(\mathbf{Y},\mathbf{X})$ is a linear function, Z is a normalization constant, and $p_j \in \{1,2\}$

Templating Language

Easy to define via **interpretable relaxation** from logical rules to hinge-loss functions using a templating language called **probabilistic soft logic** (PSL)

Example:

$$\lambda \ : \ \textsc{Label}(D_1, L) \wedge \textsc{Link}(D_1, D_2) \Rightarrow \textsc{Label}(D_2, L)$$

$$\Longleftrightarrow$$

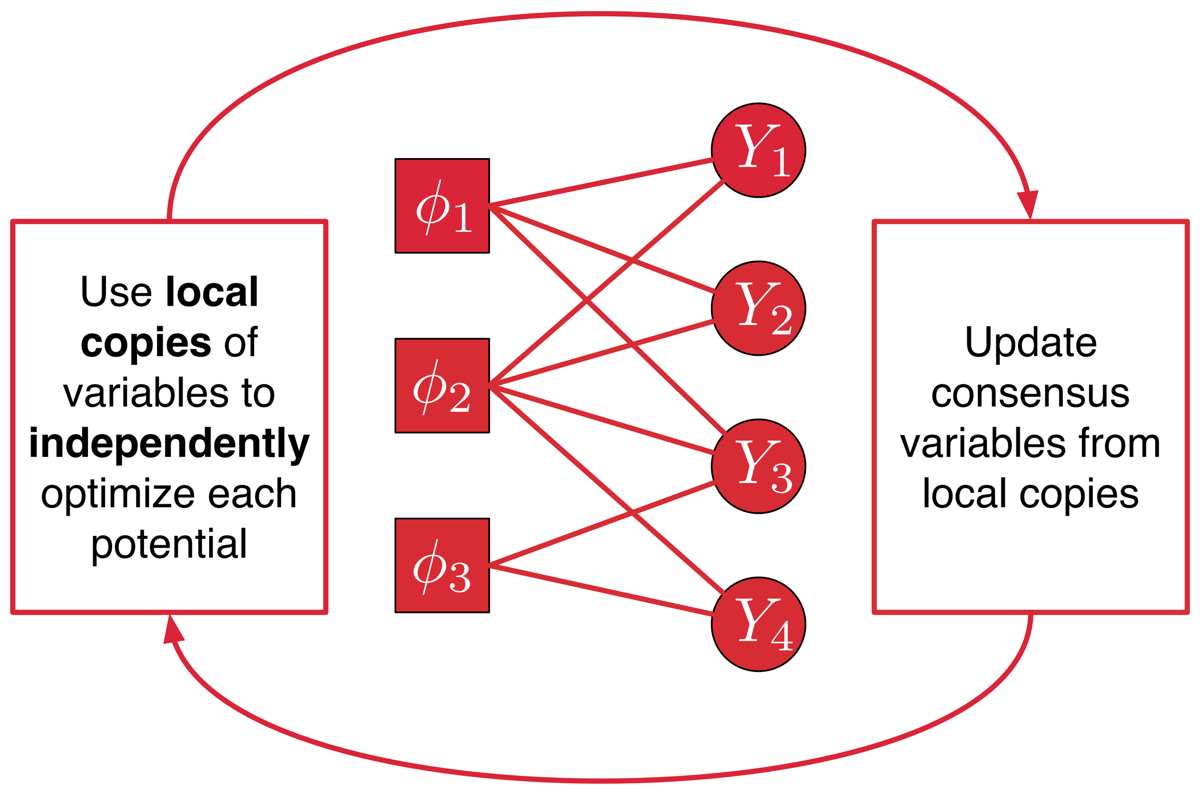
$$\lambda \cdot \max\{\textsc{Label}(D_1, L) + \textsc{Link}(D_1, D_2) - \textsc{Label}(D_2, L) - 1 \ , \ 0\}$$

Fast, Convex MPE Inference

- Hinge-loss Markov random fields are log-concave densities
- New MPE inference algorithm based on the alternating direction method of multipliers (ADMM) is highly scalable

	Citeseer	Cora	Epinions
HL-MRF-Q	0.42	0.70	0.32
HL-MRF-L	0.46	0.50	0.28
MRF	110.96	184.32	212.36

Average inference times in seconds vs. MC-SAT for discrete MRFs

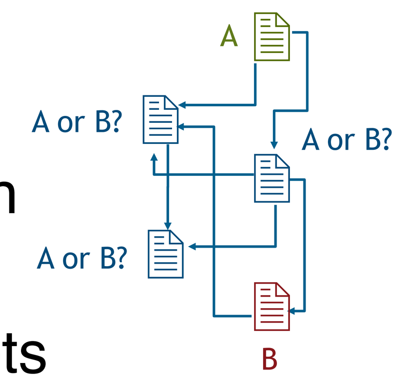


Fast Supervised Learning

- Learn tied weights Λ with
  - Approximate max likelihood: $\frac{\partial \log p(\mathbf{Y}|\mathbf{X})}{\partial \Lambda_q} = \mathbb{E}_\Lambda\left[\Phi_q(\mathbf{Y},\mathbf{X})\right] - \Phi_q(\mathbf{Y},\mathbf{X})$
  - Max pseudolikelihood: $\frac{\partial \log P^*(Y|X)}{\partial \Lambda_q} = \sum_{i=1}^{n} \mathbb{E}_{Y_i|\mathrm{MB}}\left[\sum_{j \in t_q : i \in \phi_j} \phi_j(\mathbf{Y},\mathbf{X})\right] - \Phi_j(\mathbf{Y},\mathbf{X})$
  - Large margin: $\min_{\Lambda \geq 0} \frac{1}{2}||\Lambda||^2 + C\xi \ \ \text{s.t.} \ \ \Lambda^\top(\Phi(\mathbf{Y},\mathbf{X}) - \Phi(\tilde{\mathbf{Y}},\mathbf{X})) \leq -L(\mathbf{Y},\tilde{\mathbf{Y}}) + \xi, \forall \mathbf{Y}$
- Fast inference enables fast learning

Collective Classification

- Labels in graph depend on neighbors' labels
- Learn propensity of each label value to propagate
- Citation network data sets



	Citeseer	Cora
HL-MRF-Q	**0.729**	**0.818**
HL-MRF-L	**0.729**	**0.808**
MRF	0.715	0.797

Average classification accuracy

Social-trust Prediction

- Who trusts whom in social networks?
- Easily encode **social-science theories**, such as structural balance theory, as logical rules
- Epinions data set

	ROC	P-R (+)	P-R (-)
HL-MRF-Q	**0.832**	**0.979**	**0.482**
HL-MRF-L	0.757	0.963	0.333
MRF	0.725	0.963	0.298

Average areas under curves on Epinions data set

Fast performance and state-of-the-art accuracy on four diverse tasks!

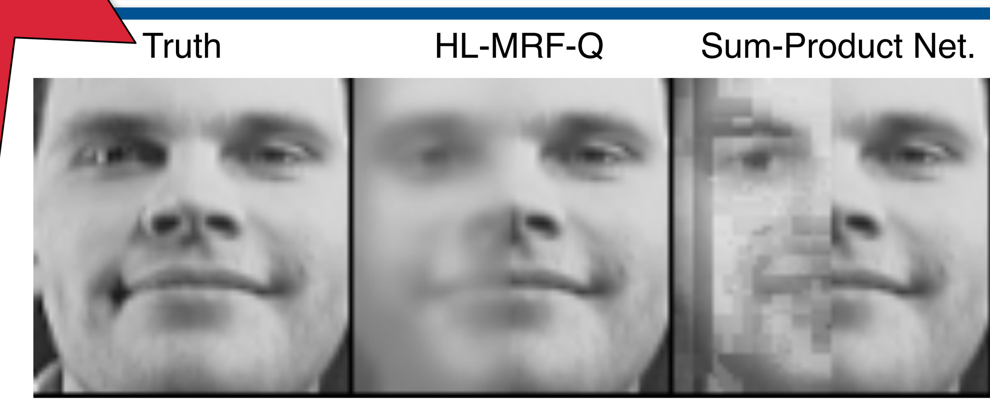
Preference Prediction

- How will a user rate something based on ratings of similar users?
- Compared to Bayesian probabilistic matrix factorization (BPMF)
- Jester jokes data set

	NMSE	NMAE
HL-MRF-Q	0.0738	0.2297
HL-MRF-L	0.0544	**0.1875**
BPMF	**0.0501**	**0.1832**

Normalized mean square and absolute errors on Jester data set

Image Reconstruction



	HL-MRF-Q	SPN
Caltech-Left	1741	1815
Caltech-Bottom	1910	1924
Olivetti-Left	927	942
Olivetti-Bottom	1226	918

Mean squared pixel error on 0-255 grayscale

This work was supported by NSF grants CCF0937094 and IIS1218488, and IARPA via DoI/NBC contract number D12PC00337. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the of- ficial policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

To get the code and learn more: http://psl.umiacs.umd.edu