**Pair**-D**ual** Learning for Fast Training of Latent Variable Hinge-Loss MRFs

Stephen H. Bach* Bert Huang* Jordan Boyd-Graber Lise Getoor
Maryland Virginia Tech Colorado UC Santa Cruz

* Equal Contributors

---

**Introduction**

- **Latent variables** are very useful in large-scale structured prediction, but they complicate learning!
- A naive approach is not scalable, and smarter methods exist for discrete models only.
- We introduce a new method for continuous MRFs that interleaves parameter and inference updates.

**Learning with Latent Variables**

- The maximum likelihood learning objective for models with latent variables contains two inner inference problems:

  \[
  \arg\min_{\mathbf{w}} \max_{\varphi \in \Delta(\mathbf{y}, \mathbf{z})} \min_{\mathbf{q} \in \Delta(\mathbf{z})} \frac{\lambda}{2} \|\mathbf{w}\|^2 - \mathbb{E}_\mathbf{y} [\mathbf{w}^\top \varphi(\mathbf{x}, \mathbf{y}, \mathbf{z})] + H(\mathbf{q}) \\
  + \mathbb{E}_\mathbf{y} [\mathbf{w}^\top \varphi(\mathbf{x}, \mathbf{y}, \mathbf{z})] - H(q)
  \]

- At a high level, this objective has a simple structure:

  \[
  \begin{align*}
  \text{Optimize } & \mathbf{w} \\
  \text{Inference in } & P(y, z|x; \mathbf{w}) \\
  \text{Inference in } & P(z|x, \hat{y}, \hat{w})
  \end{align*}
  \]

- But repeatedly performing inference is very expensive!
- Supervised learning can be sped up by interleaving inference and parameter updates, e.g., Taskar et al. [ICML 2005] and Meshi et al. [ICML 2010].
- Schwing et al. [ICML 2012] and Chen et al. [ICML 2015] propose interleaving updates for discrete latent models.
- Any new method for continuous variables must solve the problems of intractable expectations and entropies.

**Paired-Dual Learning**

- We choose point distributions for variational families and construct entropy surrogates to make objective tractable.
- Paired-dual learning replaces both inner inferences with augmented Lagrangians and optimizes jointly:

  \[
  \arg\min_{\mathbf{w}} \max_{\mathbf{y}, \mathbf{z}, \mathbf{\alpha}, \mathbf{\alpha}'} \min_{\mathbf{z}', \mathbf{\alpha}'} \max_{\mathbf{z}'} \lambda \|\mathbf{w}\|^2 - L_w(y, z, \mathbf{\alpha}, \mathbf{\hat{y}}) + L'_w(z', \mathbf{\alpha}', \mathbf{\hat{z}}')
  \]

**Evaluation**

- Paired-dual learning is:
  - Just as accurate as traditional methods
  - So fast that it often converges before others make a single parameter update

**Hinge-Loss Markov Random Fields**

- Undirected graphical models over continuous variables with hinge-loss potential functions

  \[
  P(y, z|x; \mathbf{w}) \propto \exp \left( \sum_{j=1}^m w_j \left( \max \{ \ell_j(x, y, z), 0 \} \right)^{p_j} \right)
  \]

  where \( \ell_j \) is a linear function and \( p_j \in \{1, 2\} \)

  - Generalizes Randomized algorithms for MAX SAT
  - Relaxed MAP for discrete, logical MRFs
  - Exact MAX SAT using soft logic

- Highly scalable. ADMM-based MAP inference

**Conclusions**

- Paired-dual learning overcomes the inference bottleneck associated with learning with latent variables.
- Latent variable hinge-loss MRFs are now practical for large-scale applications.
- Paired-dual learning is also applicable to other continuous models and even discrete models.

---

**Code available:** [http://psl.cs.umd.edu](http://psl.cs.umd.edu)