**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

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

where \( \ell_j(Y, X) \) 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)

\[
\lambda : \text{LABEL}(D_1, L) \land \text{LINK}(D_1, D_2) \Rightarrow \text{LABEL}(D_2, L)
\]

Example:

\[
\lambda : \max \{ \text{LABEL}(D_1, L) + \text{LINK}(D_1, D_2) - \text{LABEL}(D_2, L) - 1, 0 \}
\]

**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

**Fast Supervised Learning**

- Learn tied weights \( \Lambda \) with
- Approximate max likelihood:
  \[
  \frac{\partial \log p(Y | X)}{\partial \Lambda} = E_{\Lambda} \{ \phi(Y, X) \} - \phi(Y, X)
  \]
- Max pseudolikelihood:
  \[
  \frac{\partial \log p^*(Y | X)}{\partial \Lambda} = \sum_{i=1}^{n} E_{Y_i | \text{MB}} \left[ \sum_{c_i \in C_i} \phi(Y, X) \right] - \phi(Y, X)
  \]
- Large margin:
  \[
  \min_{\Lambda \geq 0} \frac{1}{2} ||\Lambda||^2 + C\xi \ s.t. \ \Lambda^\top (\phi(Y, X) - \phi(Y, X)) \leq -L(Y, Y) + \xi, \forall Y
  \]
- 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

<table>
<thead>
<tr>
<th></th>
<th>Citeseer</th>
<th>Cora</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL-MRF-Q</td>
<td>0.729</td>
<td>0.818</td>
</tr>
<tr>
<td>HL-MRF-L</td>
<td>0.729</td>
<td>0.808</td>
</tr>
<tr>
<td>MRF</td>
<td>0.715</td>
<td>0.797</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th></th>
<th>ROC</th>
<th>P-R (+)</th>
<th>P-R (-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL-MRF-Q</td>
<td>0.832</td>
<td>0.979</td>
<td>0.482</td>
</tr>
<tr>
<td>HL-MRF-L</td>
<td>0.757</td>
<td>0.963</td>
<td>0.333</td>
</tr>
<tr>
<td>MRF</td>
<td>0.725</td>
<td>0.963</td>
<td>0.298</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th></th>
<th>NMSE</th>
<th>NMAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL-MRF-Q</td>
<td>0.0738</td>
<td>0.2297</td>
</tr>
<tr>
<td>HL-MRF-L</td>
<td>0.0541</td>
<td>0.1875</td>
</tr>
<tr>
<td>BPMF</td>
<td>0.0501</td>
<td>0.1832</td>
</tr>
</tbody>
</table>

**Image Reconstruction**

Use local copies of variables to independently optimize each potential

\[ \phi \]

Update consensus variables from local copies

\[ \phi \]

Normalized mean square and absolute errors on Jester data set

**To get the code and learn more:** [http://psl.umiacs.umd.edu](http://psl.umiacs.umd.edu)