Learning Greedy Policies for the Easy-First Framework

Jun Xie, Chao Ma, Janardhan Rao Doppa, Prashanth Mannem, Xiaoli Fern, Tom Dietterich, Prasad Tadepalli
Oregon State University
A 4.2 magnitude earthquake struck near eastern Sonoma County.

A tremor struck in Sonoma County.
A 4.2 magnitude earthquake struck near eastern Sonoma County.

A tremor struck in Sonoma County.

1. Begin with every mention in its own cluster
A 4.2 magnitude earthquake struck near eastern Sonoma County.

A tremor struck in Sonoma County.

1. Begin with every mention in its own cluster
2. Evaluate all possible merges with a scoring function and select the highest scoring merge (easiest)
The Easy-First Framework: Example

A 4.2 magnitude earthquake struck near eastern Sonoma County.

A tremor struck in Sonoma County.

1. Begin with every mention in its own cluster
2. Evaluate all possible merges with a scoring function and select the highest scoring merge (easiest)
3. Repeat until stopping condition is met
Easy First Training

Initial State

$S_0$  

$S_1$  

$S_2$  

$S_3$  

$S_T$

Weight Update

$S_0$ Initial State

$S_1$

$S_2$

$S_3$

$S_T$

$S_0$ Initial State

$S_1$

$S_2$

$S_3$

$S_T$

Weight Update

$f(a) = 0.48$

$f(b) = 0.38$

$f(c) = 0.37$

$f(d) = 0.39$

Weight Update

$f(e) = 0.27$

$f(g) = 0.39$

$f(h) = 0.44$

$f(i) = 0.52$

Weight Update

$f(j) = 0.34$

$f(k) = 0.38$

$f(m) = 0.55$

$f(n) = 0.63$
Possible goal: learn a scoring function such that:

in every state **ALL good actions** are ranked higher than all bad actions

A better goal: learn a scoring function such that in every state **ONE good action** is ranked higher than all bad actions
Proposed Objective for Update

• Goal: find a linear function such that it ranks one good action higher than all bad actions
  – This can be achieved by a set of constraints
    \[
    \max_{g \in G} w \cdot x_g > w \cdot x_b + 1
    \]
    for all \( b \in B \)

• Our Objective:
  • Use hinge loss to capture the constraints
  • Regularization to avoid overly aggressive update

\[
\arg\min_w \frac{1}{B} \sum_{b \in B} (1 - \max_{g \in G} w \cdot x_g + w \cdot x_b)_+ + \lambda \|w - w_c\|^2
\]
Optimization

• Majorization Minimization algorithm to find a local optimal solution.

• In each MM iteration:
  – Let $x_g^*$ be the current highest scoring good action
  – Solve following convex objective (via subgradient descent)

$$\arg\min_w \frac{1}{|B|} \sum_{b \in B} \left( 1 - \max_{g \in G} w \cdot x_g^* + w \cdot x_b \right)_+ + \lambda \|w - w_c\|^2$$
Contrast with Existing Methods

- Average-good vs. average-bad (AGAB)
- Best-good vs. best-bad (BGBB)
- Proposed method: Best-good vs. violated-bad (BGVB)
Experiment I: cross-document entity and event coref

Results on EECB corpus (Lee et al., 2012)

![Bar chart showing comparison of different metrics on MUC, B-CUBE, CEAF_e, and CoNLL. The metrics include BGBB, R-BGBB, BGVB, R-BGVB, and Lee et al.]
Experiment II: within-doc Coref

Results on OntoNotes

![Bar chart showing results on OntoNotes for MUC, B-CUBE, CEAF_e, and CoNLL for BGBB, R-BGBB, BGV, and R-BGV.]
Diagnostics

• Some training statistics on ACE 2004 corpus:

<table>
<thead>
<tr>
<th>Approach</th>
<th>Total Steps</th>
<th>Mistakes</th>
<th>Recoveries</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBGVB</td>
<td>50195</td>
<td>16228</td>
<td>4255</td>
<td>0.262</td>
</tr>
</tbody>
</table>
Diagnostics

• Some training statistics on ACE 2004 corpus:

<table>
<thead>
<tr>
<th>Approach</th>
<th>Total Steps</th>
<th>Mistakes</th>
<th>Recoveries</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBGVB</td>
<td>50195</td>
<td>16228</td>
<td>4255</td>
<td>0.262</td>
</tr>
<tr>
<td>BGBB</td>
<td>50195</td>
<td>11625</td>
<td>4075</td>
<td>0.351</td>
</tr>
</tbody>
</table>

BGBB corrects errors more aggressively than RBGVB. This is a strong evidence that overfitting does happen with BGBB.
Contributions

• We precisely represent the learning goal for Easy First as an optimization problem
• We develop an efficient Majorization Minimization algorithm to optimize the proposed objective
• Achieve highly competitive results against state-of-the-art for both within- and cross-document coref