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

The Easy-First Framework: Example

A 4.2 magnitude earthquake struck near eastern Sonoma County.

Doc 1

A tremor struck in Sonoma County.

Doc 2

The Easy-First Framework: Example

A 4.2 magnitude earthquake struck near **eastern Sonoma County**.

Doc 1

A tremor struck in **Sonoma County.**

Doc 2

A 4.2 magnitude earthquake

eastern Sonoma County

A tremor

Sonoma County

1. Begin with every mention in its own cluster

The Easy-First Framework: Example

A 4.2 magnitude earthquake struck near **eastern Sonoma County**.

Doc 1

A tremor struck in **Sonoma County.**

Doc 2

A 4.2 magnitude earthquake

eastern Sonoma County

A tremor

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)

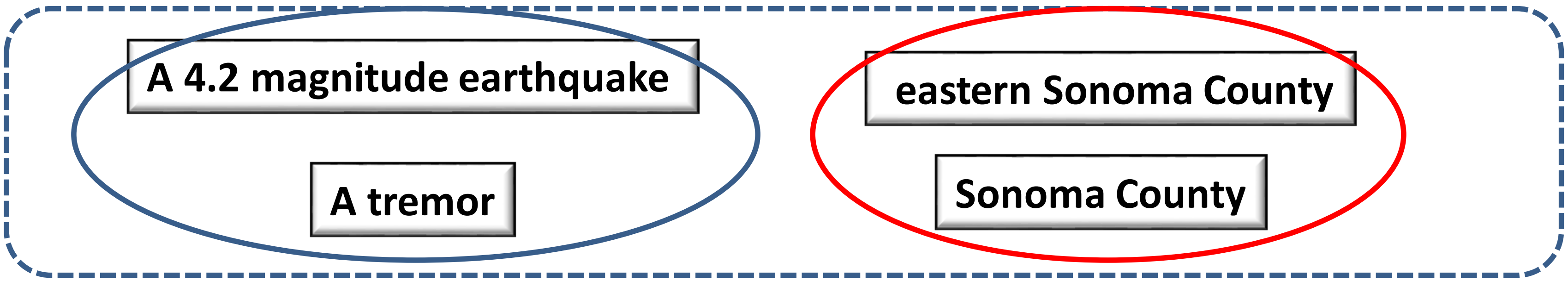
The Easy-First Framework: Example

A 4.2 magnitude earthquake struck near **eastern Sonoma County**.

Doc 1

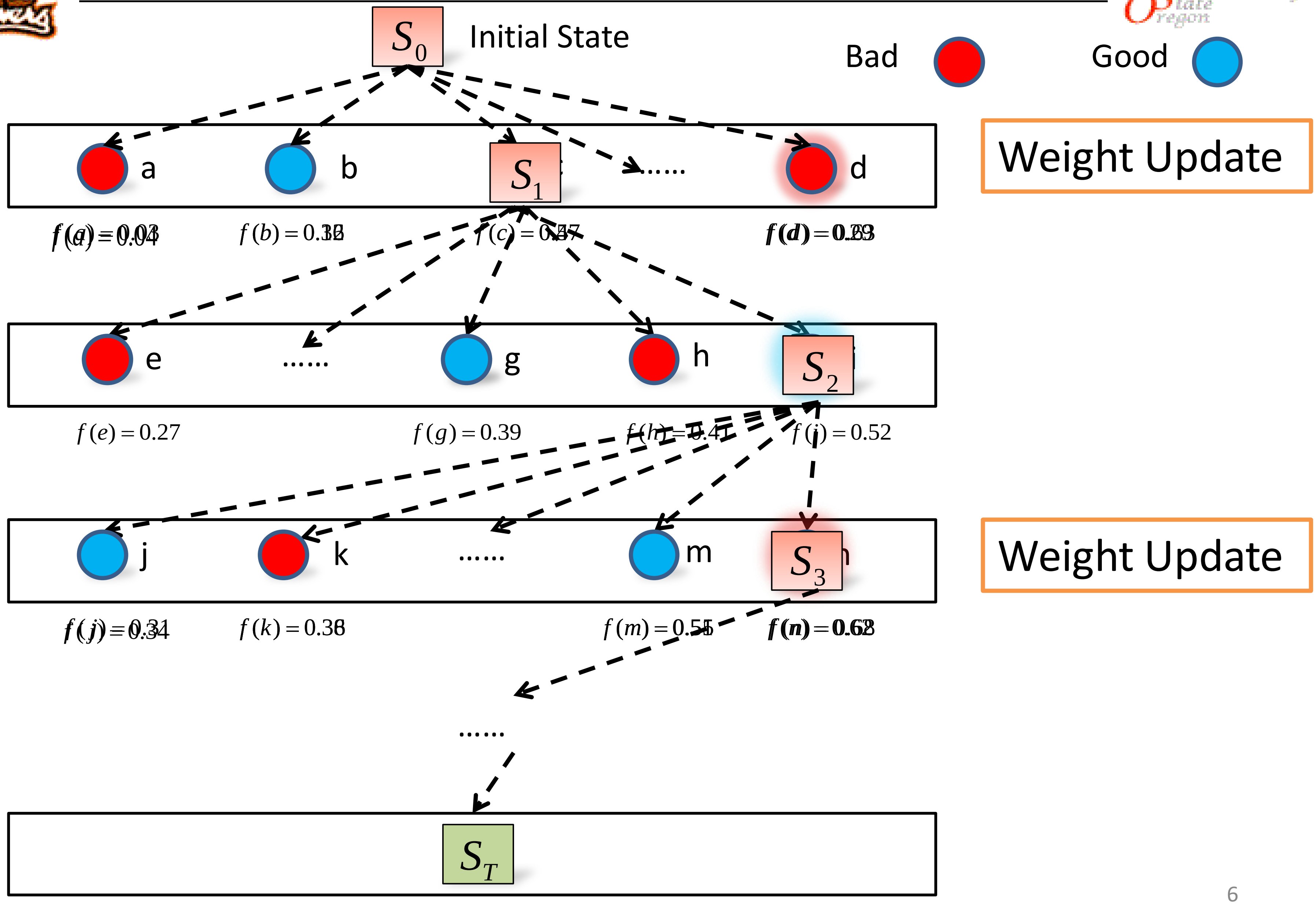
A tremor struck in **Sonoma County.**

Doc 2



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



Learning Scoring Function

Possible goal: learn a scoring function such that:

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

Over-Constrained Goal

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

$$\boxed{\operatorname*{argmin}_{w} \frac{1}{|B|} \sum_{b \in B} \left(1 - \max_{g \in G} w \cdot x_g + w \cdot x_b\right)_+} + \boxed{\lambda \left\| w - w_c \right\|^2}$$

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)

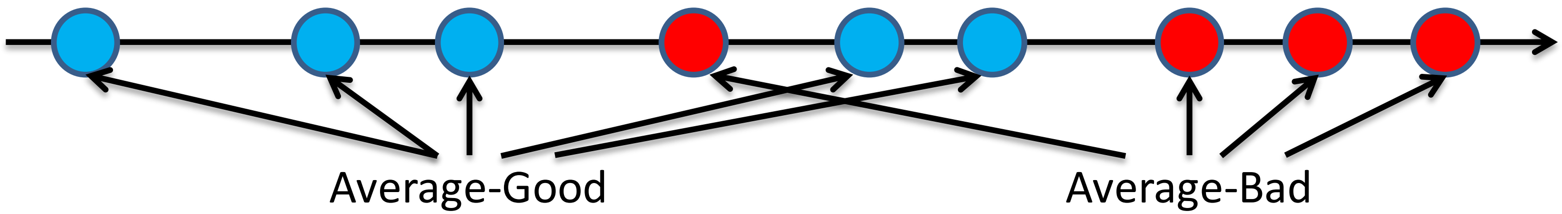
$$\underset{w}{\operatorname{argmin}} \frac{1}{|B|} \sum_{b \in B} (1 - \cancel{\max_{g \in G} w \cdot x_g} + w \cdot x_b)_+ + \lambda \| w - w_c \|^2$$

$$\boxed{w \cdot x_g^*}$$

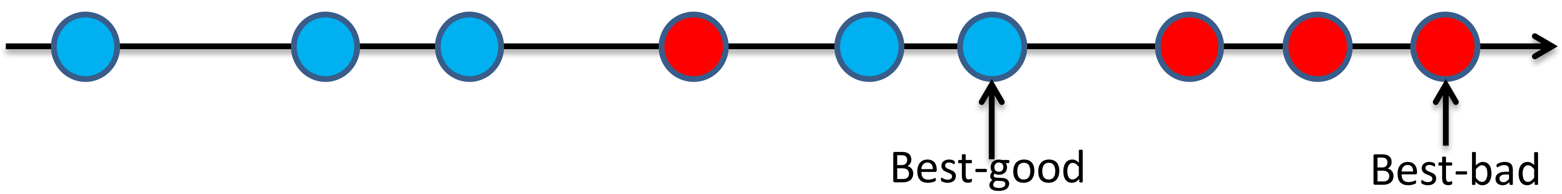
Contrast with Existing Methods

Bad ●   Good ●

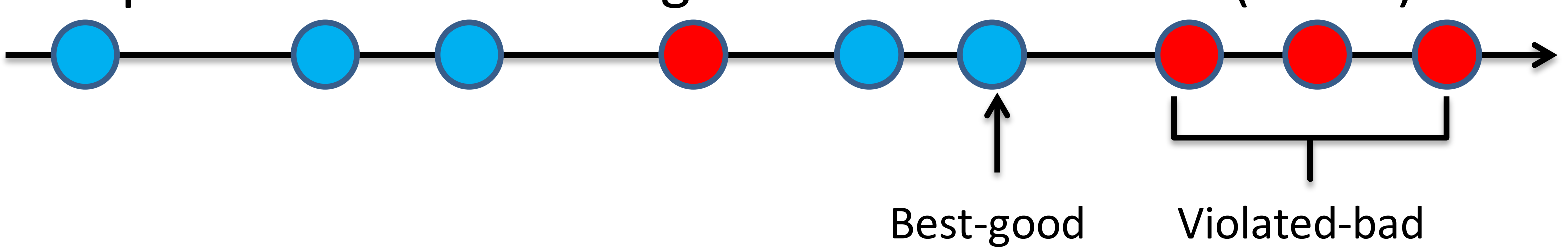
- Average-good vs. average-bad (AGAB)



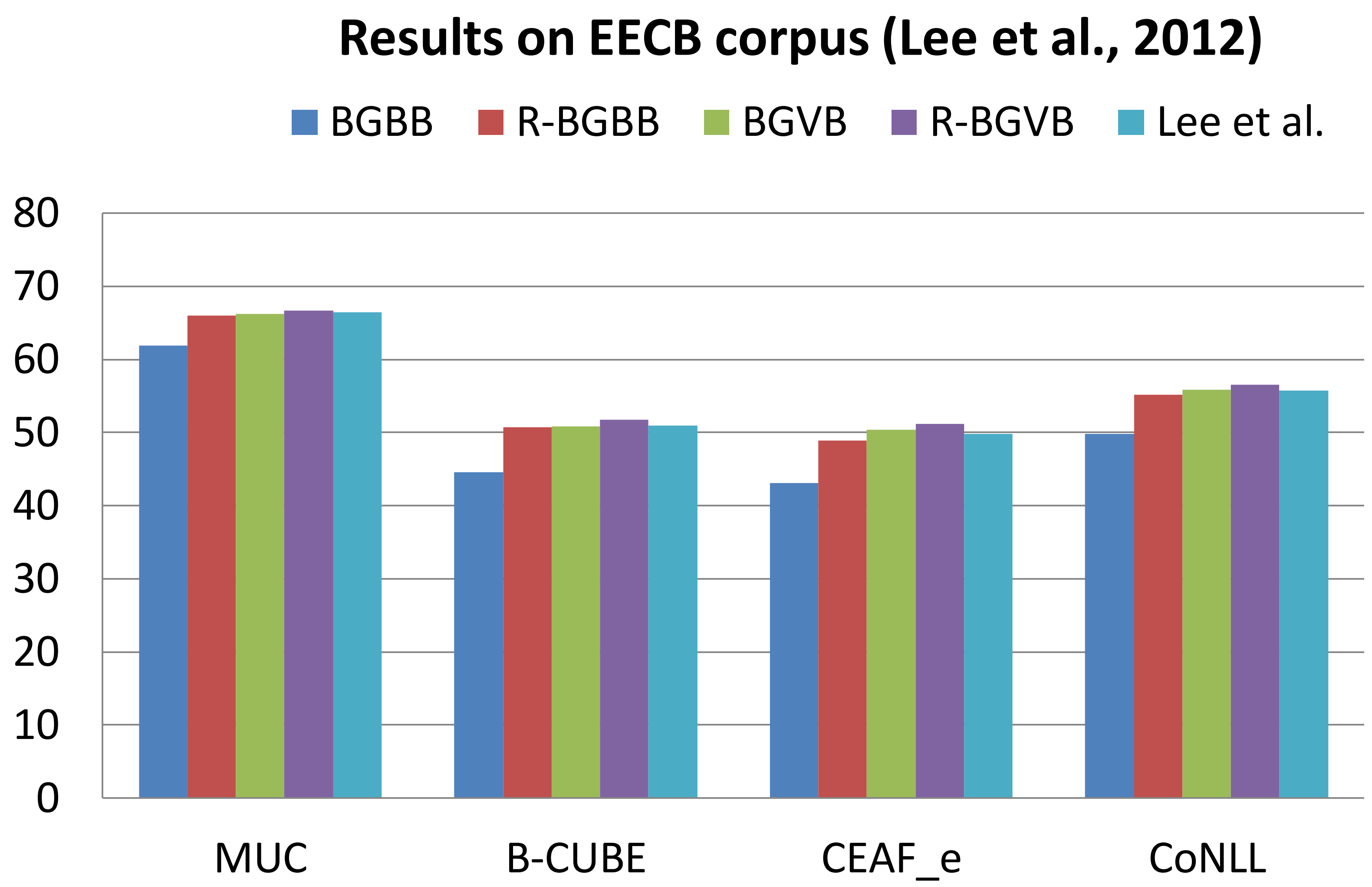
- Best-good vs. best-bad (BGBB)



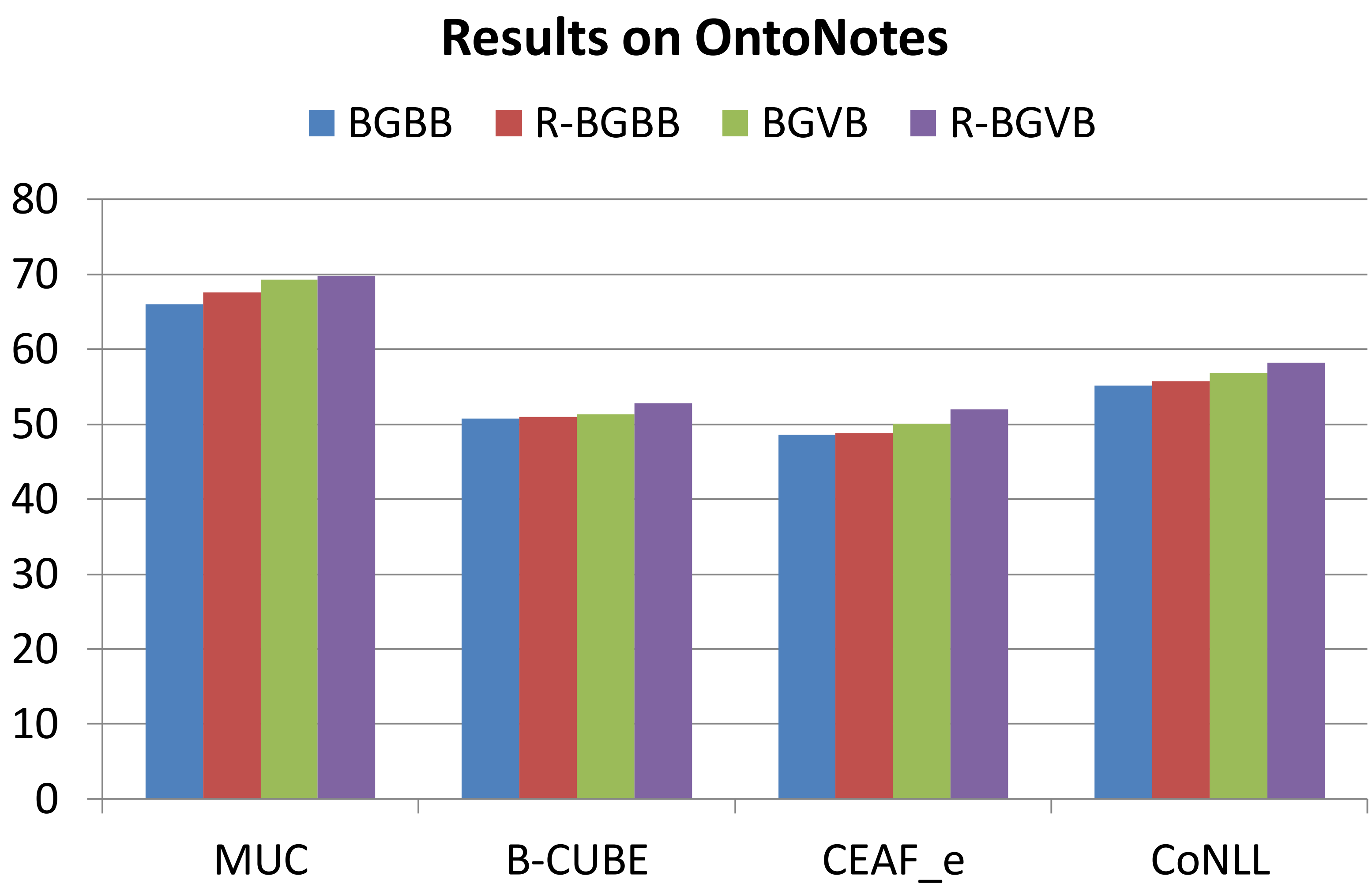
- Proposed method: Best-good vs. violated-bad (BGVB)



Experiment I: cross-document entity and event coref



Experiment II: within-doc Coref



Diagnostics

- Some training statistics on ACE 2004 corpus:

Approach	Total Steps	Mistakes	Recoveries	Percentage
RBGVB	50195	16228	4255	0.262

Diagnostics

- Some training statistics on ACE 2004 corpus:

Approach	Total Steps	Mistakes	Recoveries	**Percentage**
RBGVB	50195	16228	4255	**0.262**
BGBB	50195	11625	4075	**0.351**

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