Metric-Optimized Example Weights

Sen Zhao, Mahdi Milani Fard, Harikrishna Narasimhan, Maya Gupta

Google Research

Motivation: Building a Ranking Model

Goal: positive precision@3 globally, and not negative in any specific locales.

Training Data: Jan - Oct

Testing Data: Nov - Dec

Train with pairwise hinge loss.

Motivation: Building a Ranking Model

Attempt 1: Train a ranking model on global data.

- Good global precision@3, but negative in **Japan** and **Brazil**.

Motivation: Building a Ranking Model

Attempt 1: Train a ranking model on global data.

- Good global precision@3, but negative in **Japan** and **Brazil**.

Attempt 2: Upweight **Japan** and **Brazil** training data.

- Good metric in **Japan** and **Brazil**, but negative in **UK** and **India**.

Motivation: Building a Ranking Model

Attempt 1: Train a ranking model on global data.

- Good global precision@3, but negative in **Japan** and **Brazil**.

Attempt 2: Upweight **Japan** and **Brazil** training data.

- Good metric in **Japan** and **Brazil**, but negative in **UK** and **India**.

Attempt 3: Upweight **UK** and **India** training data.

- **US** turns negative….

Motivation: Building a Ranking Model

Attempt 1: Train a ranking model on global data.

- Good global precision@3, but negative in **Japan** and **B**

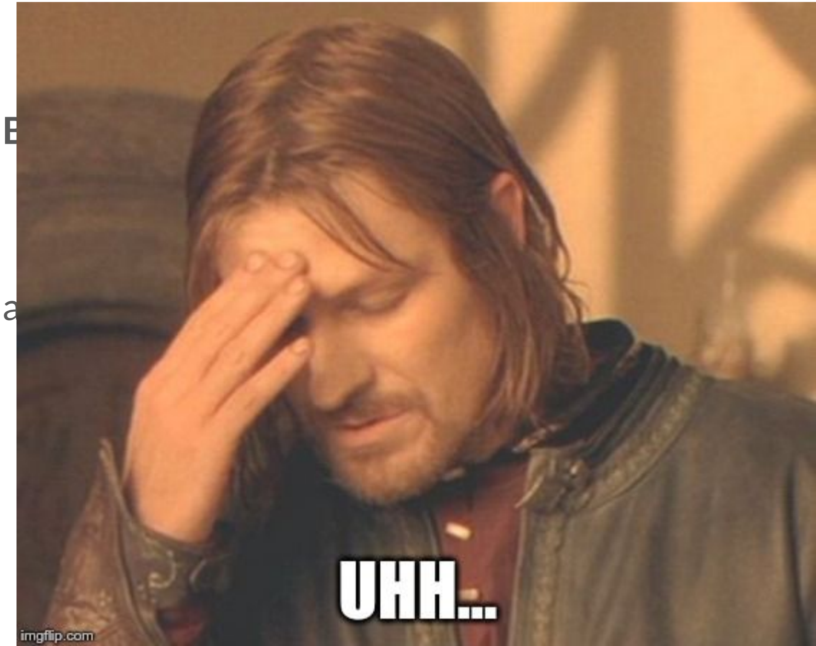
Attempt 2: Upweight **Japan** and **Brazil** training data.

- Good metric in **Japan** and **Brazil**, but negative in **UK** a

Attempt 3: Upweight **UK** and **India** training data.

- **US** turns negative….

Attempt 4: ...



A Practitioner’s Challenge

Training		Evaluation
Training Distribution		Testing Distribution
(Jan - Oct)	≠	(Holiday Season)
Training Loss		Testing Metric
(Pairwise Hinge)		(Precision@3)

Metric-Optimized Example Weights (MOEW)

MOEW learns the optimal weighting on training examples to **maximize the testing metric**.

- Suitable for ***any*** loss and ***any*** (black-box, non-differentiable) metrics.
- Accompanied by theoretical analysis (generalization bounds etc.).

Formulation

The main model **θ** is an ERM problem with weighted loss:

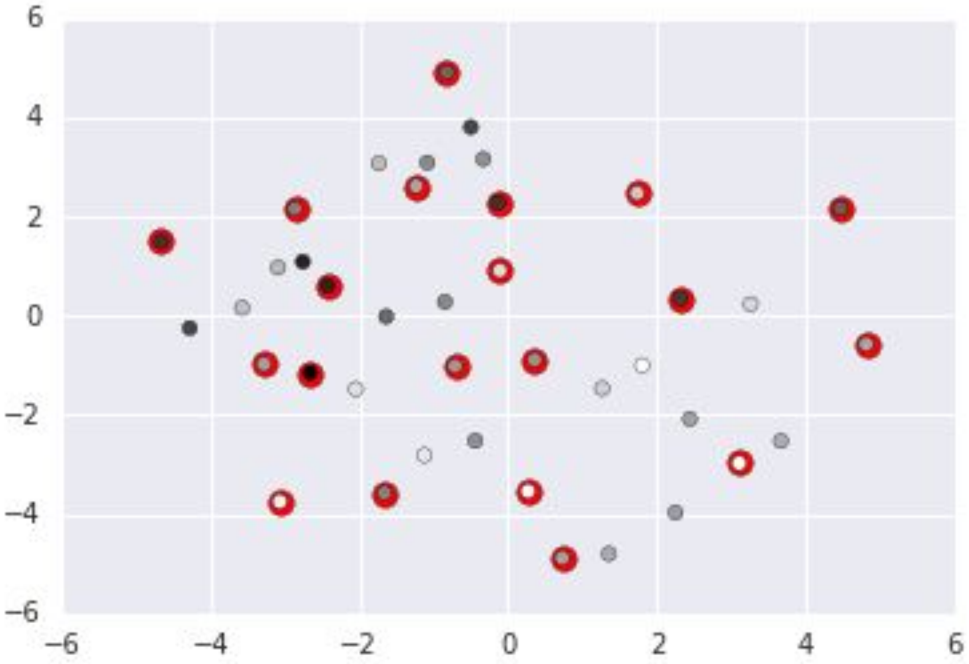
$$\hat{\theta}(\alpha) = \arg\min_{\theta} \sum_{j \in \mathcal{T}} w\left(x_j, y_j; \alpha\right) L\left(h\left(x_j; \theta\right), y_j\right)$$

The weighting model **ω** has one parameter **α** that is learned to maximize validation metric:

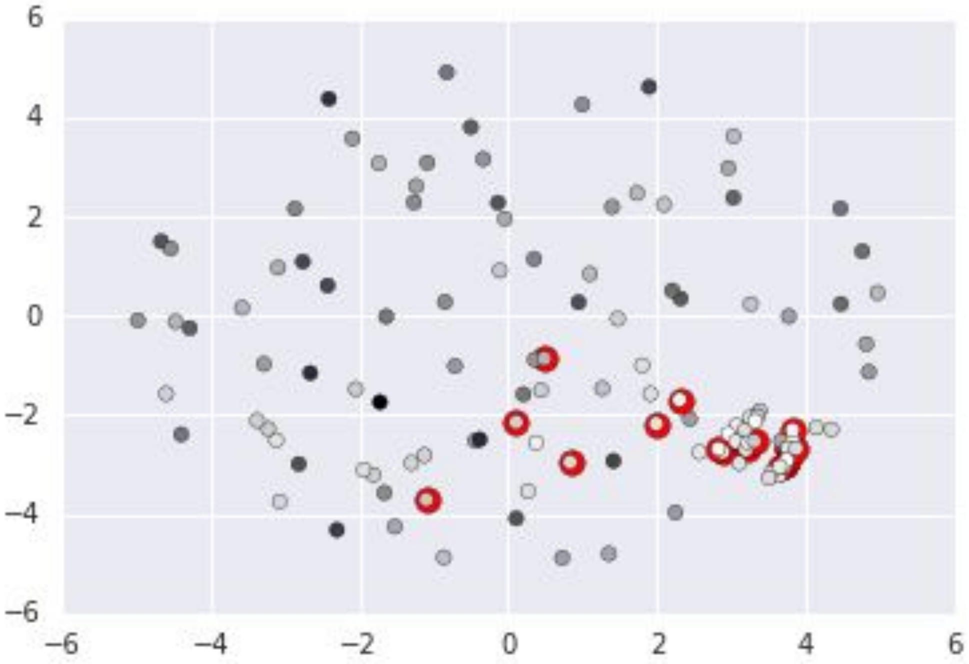
$$\hat{\alpha} = \arg\max_{\alpha} \hat{M}(x_{\mathcal{V}}, y_{\mathcal{V}}; \hat{\theta}(\alpha))$$

Iteratively optimize...

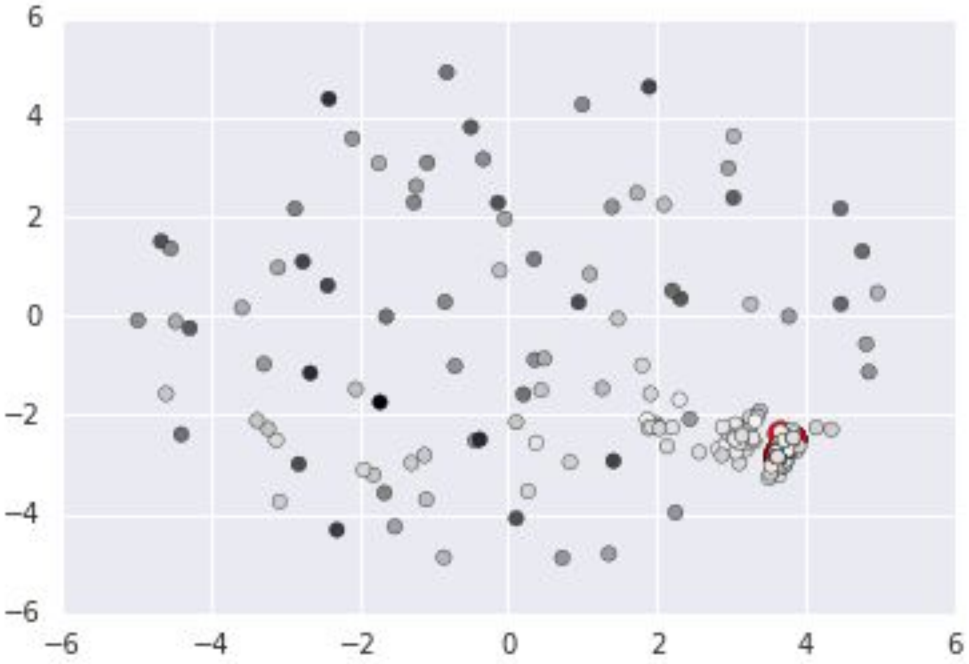
A Sneak Peek of MOEW



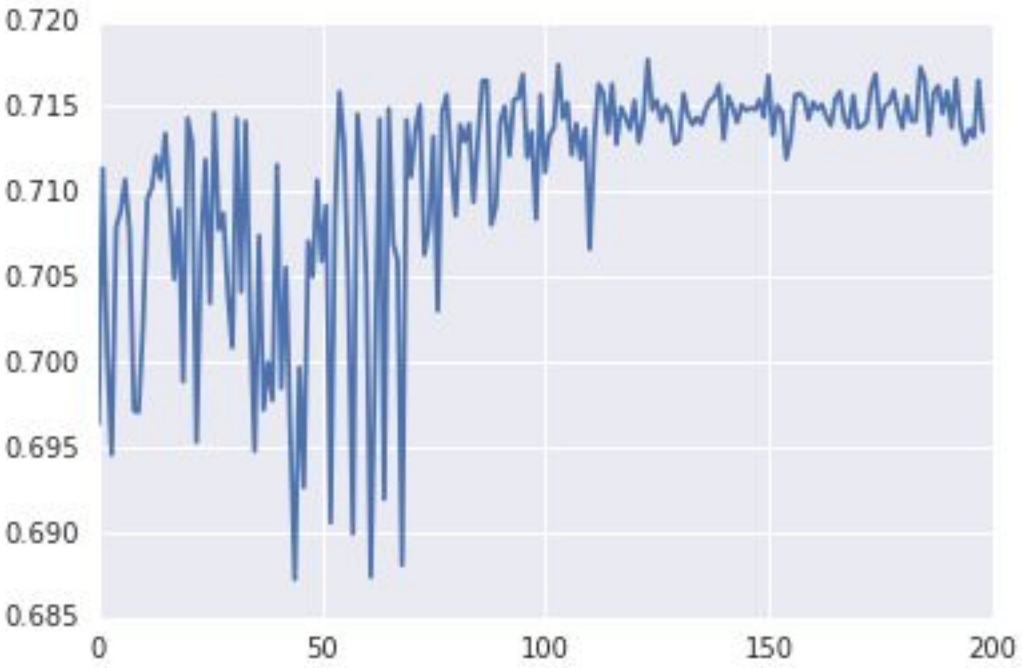
A Sneak Peek of MOEW



A Sneak Peek of MOEW



A Sneak Peek of MOEW



Poster

Tonight 06:30 -- 09:00 PM @ Pacific Ballroom #122