http://linqs.cs.umd.edu

Large-margin Structured Learning for Link Ranking

Stephen H. Bach[1], Bert Huang[1], and Lise Getoor[2]
[1] University of Maryland, College Park
[2] University of California, Santa Cruz

Introduction

- **Link prediction** is the task of predicting which nodes are linked in a network
- We introduce a **large-margin method** for learning to **rank in structured domains**, where ranking positions are interdependent
- By training a structured predictor to rank, we **improve performance** on link prediction as measured by a ranking loss

Large-margin Structured Learning

With observations $\mathbf{X}$ and training output $\mathbf{Y}$ in structured output space $\mathcal{Y}$, find parameters λ of a function f_λ such that the following condition holds:

$$f_\lambda(\mathbf{Y}, \mathbf{X}) \geq f_\lambda(\tilde{\mathbf{Y}}, \mathbf{X}) + L(\mathbf{Y}, \tilde{\mathbf{Y}}), \quad \forall \tilde{\mathbf{Y}} \in \mathcal{Y}$$

which leads to this learning objective:

$$\min_{\lambda} \quad \frac{1}{2}\|\lambda\|^2 + C \max_{\tilde{\mathbf{Y}} \in \mathcal{Y}} \left(f_\lambda(\tilde{\mathbf{Y}}, \mathbf{X}) - f_\lambda(\mathbf{Y}, \mathbf{X}) + L(\mathbf{Y}, \tilde{\mathbf{Y}}) \right)$$

Perform subgradient descent, taking steps via

$$\nabla_\lambda = \lambda + C\left(\phi(\mathbf{Y}^\star, \mathbf{X}) - \phi(\mathbf{Y}, \mathbf{X})\right)$$

where

$$\mathbf{Y}^\star = \arg\max_{\tilde{\mathbf{Y}}} f_\lambda(\tilde{\mathbf{Y}}, \mathbf{X}) + L(\mathbf{Y}, \tilde{\mathbf{Y}})$$

Computational challenge is finding $\mathbf{Y}^\star$ efficiently, which depends on forms of both f_λ and L

Structured Ranking

Train a structured predictor to rank via large-margin learning, i.e., use ranking loss for $L(\mathbf{Y}, \tilde{\mathbf{Y}})$

ROC loss (one minus area under ROC curve):

$$L^{\mathrm{ROC}}(\mathbf{Y}, \tilde{\mathbf{Y}}) \equiv \frac{1}{|\mathcal{P}||\mathcal{N}|} \sum_{(i,j)|Y_i > Y_j} \mathbb{I}\left[\tilde{Y}_j > \tilde{Y}_i\right]$$

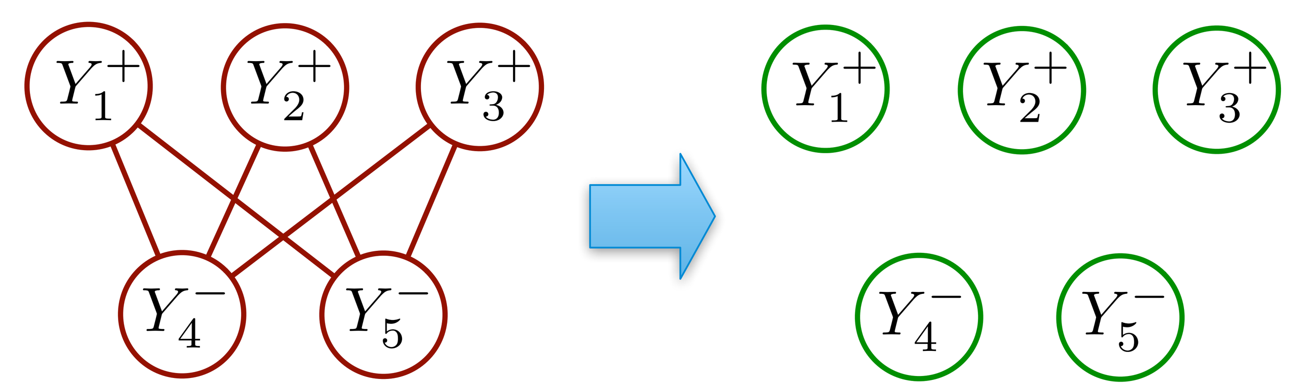
Pseudo ROC loss (convex lower bound):

$$L^{\mathrm{pROC}}(\mathbf{Y}, \tilde{\mathbf{Y}}) \equiv \frac{1}{|\mathcal{P}||\mathcal{N}|} \sum_{(i,j)|Y_i > Y_j} \max\left\{0, \tilde{Y}_j - \tilde{Y}_i\right\}$$

Gradient Computation

Finding $\mathbf{Y}^\star$ with L^{pROC} is a non-convex objective. We iteratively solve a linear approximation to find a local optimum and reduce problem size.

Approximate $|\mathcal{P}||\mathcal{N}|$ terms with $|\mathcal{P}| + |\mathcal{N}|$ terms



where $\mathcal{P}$ is set of true links and $\mathcal{N}$ is false links

Algorithm 1 Gradient Computation for L^{pROC}

Input: model f_λ, input $\mathbf{X}$, training output $\mathbf{Y} = \mathcal{P} + \mathcal{N}$, initial guess $\tilde{\mathbf{Y}}$

Output: $\mathbf{Y}^\star = \arg\max_{\tilde{\mathbf{Y}}} f_\lambda(\tilde{\mathbf{Y}}, \mathbf{X}) + L^{\mathrm{pROC}}(\mathbf{Y}, \tilde{\mathbf{Y}})$

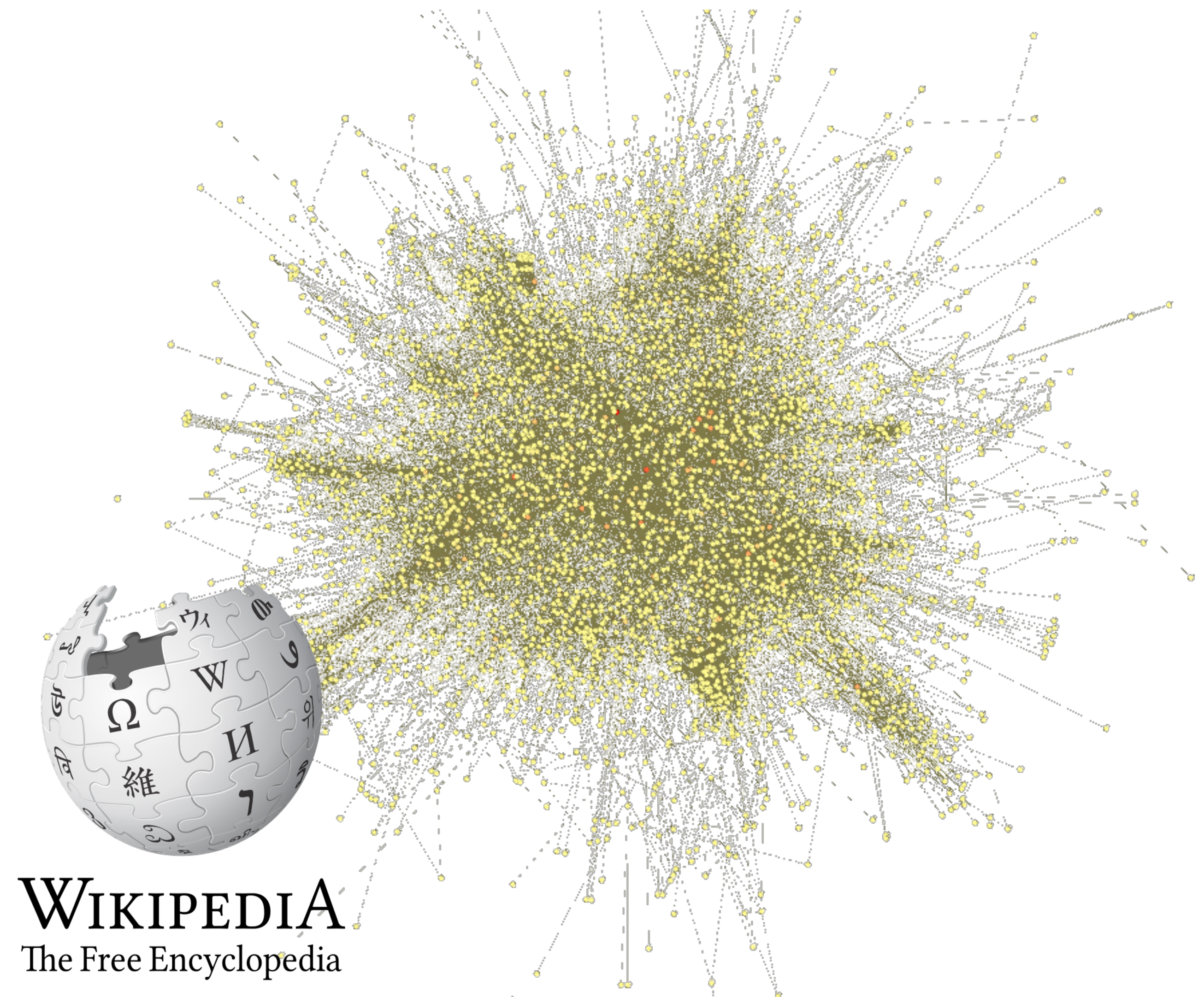
while not converged **do**
  $\tilde{\mathbf{Y}} \leftarrow \mathtt{Sort}(\tilde{\mathbf{Y}})$
  $\mathbf{c} \leftarrow \mathtt{UpdateCounts}(\tilde{\mathbf{Y}})$
  $\tilde{\mathbf{Y}} \leftarrow \arg\max_{\tilde{\mathbf{Y}}} f_\lambda(\tilde{\mathbf{Y}}, \mathbf{X}) + \frac{1}{|\mathcal{P}||\mathcal{N}|} \sum_{i=1}^{n} c_i \tilde{Y}_i$
end while
$\mathbf{Y}^\star \leftarrow \tilde{\mathbf{Y}}$

Learn more: http://psl.cs.umd.edu

Preliminary Experimental Results

Task

Predict hyperlinks in set of featured Wikipedia articles using text of articles, category of articles, and structural dependencies among links.



Structured Predictor

Hinge-loss Markov random fields [Bach et al., UAI 2013] are **scalable** graphical models over **continuous** random variables

Most probable assignment to continuous variables can be used as a ranking of possible hyperlinks

Results

	ROC	P-R	Acc.
pROC	**0.869 (0.046)**	**0.601 (0.106)**	0.804 (0.112)
L1	0.843 (0.047)	0.556 (0.111)	**0.852 (0.086)**
Perceptron	0.842 (0.049)	0.524 (0.105)	0.667 (0.145)

Table 1: Average area under ROC and precision-recall curves and 0-1 accuracy with different loss functions during large-margin learning. Standard deviations are listed in parentheses. Scores statistically equivalent to the best scoring method by metric are typed in bold.

This work was supported by NSF grant CCF0937094 and IARPA via DoI/NBC contract number D12PC00337. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.