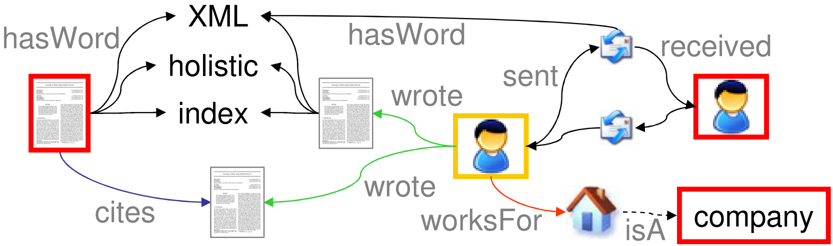
Learning Random Walks to Rank Nodes in Graphs

Alekh Agarwal   Soumen Chakrabarti
IIT Bombay

http://www.cse.iitb.ac.in/~soumen/doc/netrank

Learning to Rank

- Web search, recommender systems, database search, ...
- Vector Spaces: Learn a good scoring function $\langle \beta, \psi(x) \rangle$ from training preferences for fixed feature mapping ψ on instances x
- Graphs: Of great recent interest for modelling relationships



Need to exploit/respect information from links

Two Roles of Edges in Graph Ranking

Associative Networks

- Edges encode *similarity*
- Preference for *smooth* scoring functions
- Typically edge weights indicate extent of similarity

Random Walk Approach

- Edges indicate *endorsement*
- Motivated by Pagerank, widely used
- Each edge (u, v) has transition probability $Q(v, u) = \Pr(v|u)$

Note : Equivalent for undirected graphs.

- Typically Q fixed by hand and $\pi = Q\pi$ found
- We have an "inverse problem": given properties of π, find a transition probability matrix

Training and Testing

- Fixed directed graph G
- Training set of some node-pair preferences:

$$\text{``}u \prec v\text{''} \quad \text{means we want} \quad \text{score}(u) < \text{score}(v)$$

- More node-pair preferences in test set
- Sampling distribution over node pairs not necessarily uniform, but same for training and testing
- Transductive, in the sense that node-pair space is finite
- Performance measured by number of incorrect test set predictions
- Learner must assign scores to satisfy training pairs without overfitting

Ranking Using Random Walk (Agarwal+ 2006)

- Pagerank vector $\pi = Q\pi$, $\Pi = \mathrm{diag}(\pi)$, induces flow $q_{uv} = \pi(u) Q_{vu}$ along graph edges
- Goal: learn flow p close to q that satisfies training preferences

$$\min_{\substack{\{0 \le p_{uv}\} \\ \{0 \le s_{uv}\}}} \quad \sum_{(u,v)\in E} p_{uv} \log \frac{p_{uv}}{q_{uv}} + B \sum_{u \prec v} s_{uv} \qquad \text{(KL)}$$

$$\text{s.t.:} \quad \sum_{(u,v)\in E} p_{uv} = 1 \qquad \text{(Sum)}$$

$$\forall v \in V: \quad \sum_{(u,v)\in E} p_{uv} - \sum_{(v,w)\in E} p_{vw} = 0 \qquad \text{(Balance)}$$

$$\forall u \prec v: \quad \sum_{(w,u)\in E} p_{wu} - \sum_{(w,v)\in E} p_{wv} - s_{uv} \le 0 \qquad \text{(Pref)}$$

Limitations of Markovian Flow Approach

- Only intuitive motivation for use of q
- No known generalization bounds
- No margin in training constraints

Ranking in Associative Networks (Agarwal 2006)

- Directed Graph Laplacian $L = \mathbb{I} - \left(\frac{\Pi^{1/2} Q \Pi^{-1/2} + \Pi^{-1/2} Q \Pi^{1/2}}{2}\right)$.
- $f^\top L f = \sum_{(u,v)\in E} \pi(u) \hat{Q}_{uv} \left(\frac{f(u)}{\sqrt{\pi(u)}} - \frac{f(v)}{\sqrt{\pi(v)}}\right)^2$ enforces *smoothness*
- Scoring algorithm:

$$\min_{\substack{f:V\to\mathbb{R} \\ s=\{s_{uv}\geq 0: u \prec v\}}} \quad \frac{1}{2} f^\top L f + B \sum_{u \prec v} s_{uv} \quad \text{subject to} \tag{Lap}$$

$$f_v - f_u \geq 1 - s_{uv} \quad \forall u \prec v$$

- Generalization proved using algorithmic stability (Bousquet+ 2002)
- $f(v) \propto \sqrt{\pi(v)}$ minimizes $f^\top L f$
- I.e. prefers pagerank ordering in absence of training data

Limitations of the Laplacian Approach

- "Link as similarity hint" not universal view
- Millions of obscure pages u link to $v =$ `http://kernel-machines.org`, with $\text{score}(u) \ll \text{score}(v)$
- Score f_u can be arbitrary, even negative
- No intuitive meaning like probability as for $\pi(u)$
- Generalization depends on $\kappa = \max_{u \in V} L^+(u, u)$, hard to interpret
- Typical QP solvers compute and store large, dense L^+ in RAM

Our Contributions

- Relating Laplacian regularization with KL regularization
- Stability-based generalization bounds for random walk ranking
- Key parameters that affect generalization
- Incorporation of margin in random walk ranking
- Cost-sensitive ranking framework

Laplacian-KL Correspondence

- Algorithm(KL) returns flow $\{p_{uv}\}$
- Define node scores using flow $\{p_{uv}\}$:

$$f_p(u) = \sqrt{\sum_{w:(w,u)\in E} p_{wu}}$$

- Same rank order as Pagerank node score: $\sum_{w:(w,u)\in E} p_{wu}$
- We show that

$$\mathrm{KL}(p\|q) \le \epsilon \quad \Rightarrow \quad f_p^\top L f_p \le 4(2\epsilon \ln 2)^2$$

$\therefore$ If (KL) achieves a small KL distance, we can find scores with low Laplacian roughness penalty too

- First hint that $\mathrm{KL}(\cdot\|q)$ is a good regularizer

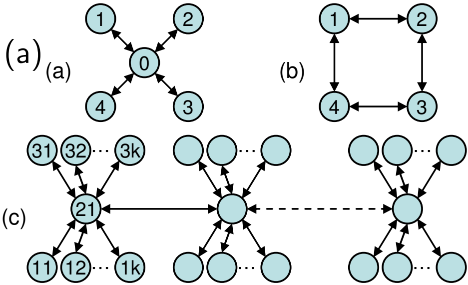
Generalization Bound for (KL)

- For any $m \geq 1$, any $\delta \in (0,1)$, the following holds w.p. $\geq 1-\delta$ over random draws of sample $\prec$ of size m:

$$R \leq R_{\mathrm{emp}} + 2\beta + (4m\beta + 1)\sqrt{\frac{\ln(1/\delta)}{2m}}$$

- Graph-agnostic case: $\beta = \frac{2\ln 2}{\lambda m}$
- G may not reduce function class (a)
- Degree bound D not enough (c)
- But G can be useful too (b)
- Key parameter: eccentricity

$$\rho = \max_{u \in V} \frac{\max_{v:(u,v)\in E} p_{uv}}{\min_{v:(u,v)\in E} p_{uv}}$$



- Modified β worsens with increasing D, ρ
- Together, ρ and D control influence of any single node

Loss Functions and Problem with Margin

- Thus far, our loss function has been

$$\ell_0(f,u,v) = \begin{cases} 0, & f(u) - f(v) < 0 \\ f(u) - f(v), & 0 \leq f(u) - f(v) \end{cases}$$

- Ideally, we want an upper bound on $0/1$ loss e.g., hinge loss

$$\ell_1(f,u,v) = \begin{cases} 0, & f(u) - f(v) < -1 \\ 1 + f(u) - f(v), & -1 \leq f(u) - f(v) \end{cases}$$

- In (KL), $\sum_{(u,v)} p_{uv} = 1$ means all node scores in $[0, 1]$
- Typically most nodes scores are very small
- Arbitrary additive margin (like "1") unattainable except for "meaningless" slacks
- Let flows $\{p_{uv}\}$ float to variable scale: $\sum_{\{uv \in E\}} p_{uv} = F$
- Luckily, $\mathrm{KL}(p\|q)$ well-defined even if p not normalized

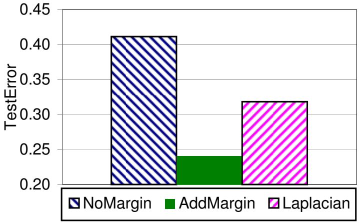
(KL) with Additive Margin

$$\min_{\substack{\{p_{uv}\},\{s_{uv}\}\\F\geq 1}} \sum_{(u,v)\in E'} p_{uv} \log \frac{p_{uv}}{q_{uv}} + B \sum_{u \prec v} s_{uv} + B_1 F^2$$

$$\text{subject to} \quad 1 + \sum_{(w,u)\in E} p_{wu} - \sum_{(w,v)\in E} p_{wv} - s_{uv} \leq 0 \quad \forall u \prec v$$

- Small F enforces large margin
- Can show (polynomial in δ) generalization bound
- Generalization worsens as upper bound on F increases

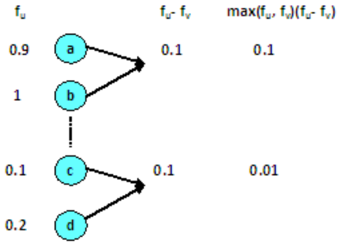
Experimental Results



- Evaluated on real and synthetic social networks
- First computed reference (unweighted) pagerank
- Secretely perturbed conductance of some edges
- Computed perturbed pagerank, considered to be the "true" hidden score
- Sampled training and test pairs from agreements and disagreements between the two rankings
- Additive margin in (KL) helps, gives best results

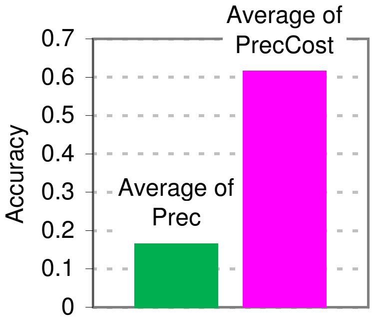
Cost Sensitivity in Ranking

- Tester wants no mistake at top of ranked list
- Excessive cognitive burden on trainer to provide total orders or "true" scores
- *Main Intuition*: Use score estimate as surrogate
- High confidence in nodes predicted to be ranked high.



- With high probability, loss is small for pairs with large f_u or f_v
- Generalization bounds follow from stability wrt g.

Cost-sensitive ranking experiments



- Number of violations in test set no longer appropriate, need cost-sensitive performance measure
- Use precision at rank k
- Cost-sensitive formulation does better than cost-ignorant counterpart
- Better for other accuracy measures too, like Kendall's tau

Summary

- Connections between Laplacian and random walk setups
- Generalization bounds in terms of intuitive graph parameters for random walk ranking
- Margin in random walk ranking, beats Laplacian approach in experiments
- Optimization more scalable for random walk approach
- A general cost-sensitive ranking framework
- Effective experimental results in modified Laplacian framework

References I

A. Agarwal, S. Chakrabarti, and S. Aggarwal, "Learning to rank networked entities," in *SIGKDD Conference*, 2006, pp. 14–23.
http://www.cse.iitb.ac.in/~soumen/doc/netrank

S. Agarwal, "Ranking on graph data," in *ICML*, 2006, pp. 25–32.
http://web.mit.edu/shivani/www/Papers/2006/icml06-graph-ranking.pdf

O. Bousquet and A. Elisseeff, "Stability and generalization," *Journal of Machine Learning Research*, vol. 2, pp. 499–526, 2002.
http://www.cmap.polytechnique.fr/~bousquet/papers/BouEli01_stability_jmlr.ps