Effective Latent Space Graph-based Re-ranking Model with Global Consistency

Hongbo Deng, Michael R. Lyu and Irwin King

Department of Computer Science and Engineering
The Chinese University of Hong Kong

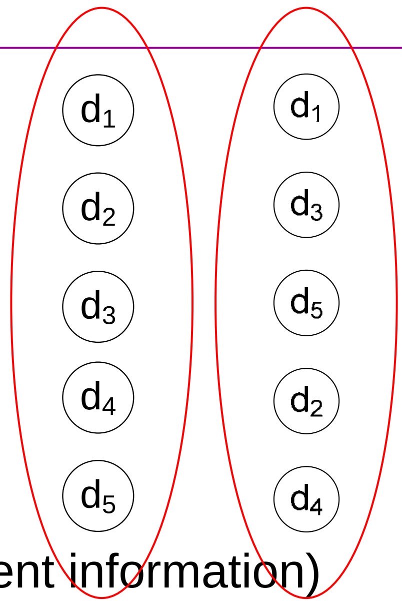
Feb. 12, 2009

Outline

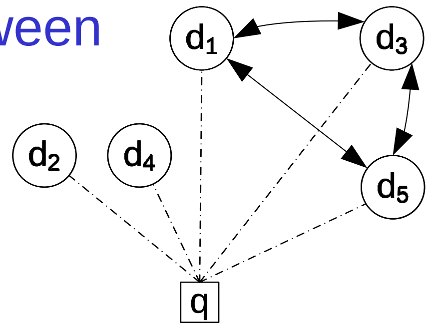
- Introduction
- Related work
- Methodology
  - Graph-based re-ranking model
  - Learning a latent space graph
  - A case study and the overall algorithm
- Experiments
- Conclusions and Future Work

Introduction

- Problem definition
  - Given a set of documents D
  - A term vector $d_i = x_i$
    - Relevance scores using VSM or LM
  - A connected graph
    - Explicit link (e.g., hyperlinks)
    - Implicit link (e.g., inferred from the content information)
  - Many other features



  - How to leverage the interconnection between documents/entities to improve the ranking of retrieved results with respect to the query?



Hongbo Deng, Michael R. Lyu and Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong

Introduction

- Initial ranking scores: relevance $q \longleftrightarrow d_n$
- Graph structure: centrality (importance, authority)
- Simple method: Combine those two parts linearly
- Limitations:
  - Do not make full use of the information
  - Treat each of them individually

- What we have done?
  - Propose a joint regularization framework
  - Combine the content with link information in a latent space graph

Related work

Structural re-ranking model

Regularization framework

Learning a latent space

- Using some variations of PageRank and HITS
  - Centrality within graphs (Kurland and Lee, SIGIR’05 & SIGIR’ 06)
  - Improve Web search results using affinity graph (Zhang et al., SIGIR’05)
  - Improve an initial ranking by random walk in entity-relation networks (Minkov et al., SIGIR’06)

Linear combination, treat the content and link individually

Related work

Structural re-ranking model

Regularization framework

Learning a latent space

- Regularization framework
  - Graph Laplacians for label propagation (two classes) (Zhu et al., ICML’03, Zhou et al., NIPS’03)
  - Extent the graph harmonic function to multiple classes (Mei et al., WWW’08)
  - Score regularization to adjust ad-hoc retrieval scores (Diaz, CIKM’05)
  - Enhance learning to rank with parameterized regularization models (Qin et al., WWW’08)

Query-independent settings

Do not consider multiple relationships between objects.

Related work

Structural re-ranking model

Regularization framework

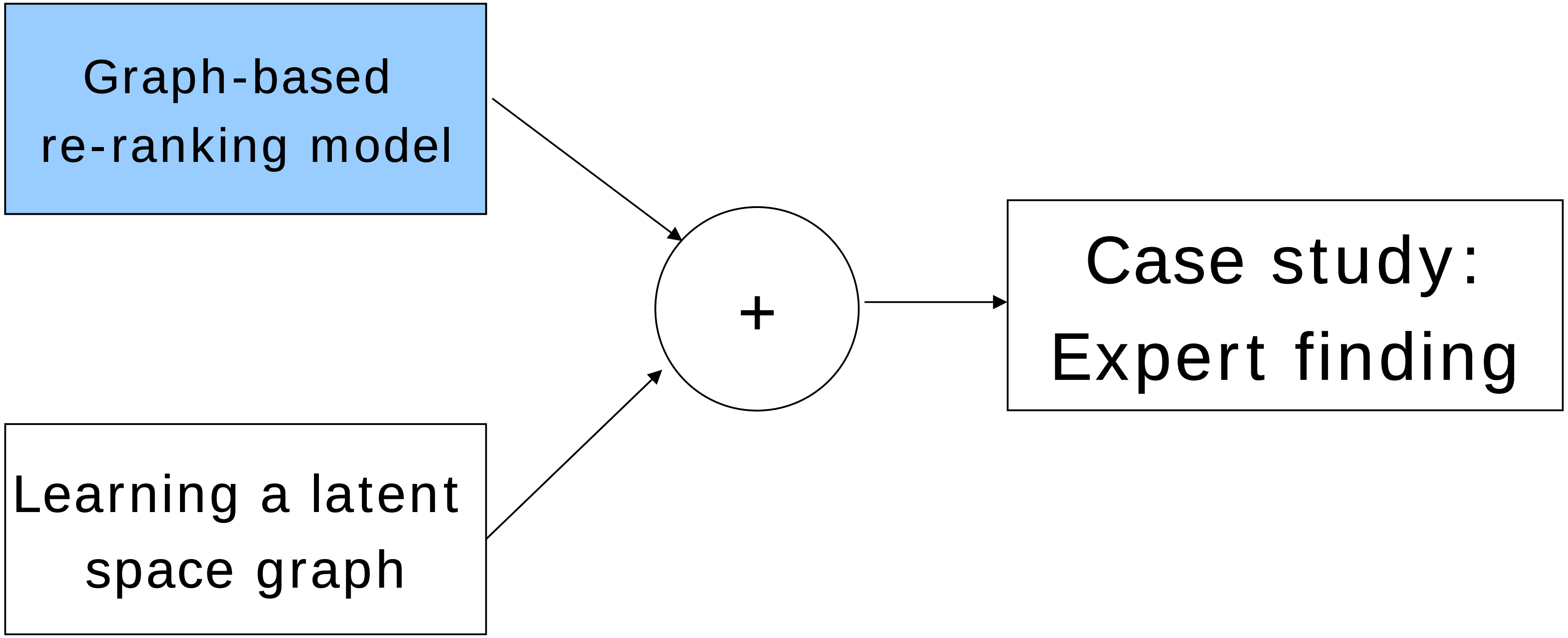
Learning a latent space

- Learning a latent space
  - Latent Semantic Analysis (LSA) (Deerwester et al., JASIS'90)
  - Probabilistic LSI (pLSI) (Hofmann, SIGIR'99)
  - pLSI + PHITS (Cohn and Hofmann, NIPS'00)
  - Combine content and link for classification using matrix factorization (Zhu et al., SIGIR'07)

Use the joint factorization to learning the latent feature.

Difference: leverage the latent feature for building a latent space graph.

Methodology



Hongbo Deng, Michael R. Lyu and Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong

Graph-based re-ranking model

- Intuition:
  - Global consistency: similar documents are most likely to have similar ranking scores with respect to a query.
  - The initial ranking scores provides invaluable information
- Regularization framework

$$R(F,q,G)=\underbrace{\frac{1}{2}\sum_{i,j=1}^{n} w_{ij}\left\|\frac{f(d_i,q)}{\sqrt{D_{ii}}}-\frac{f(d_j,q)}{\sqrt{D_{jj}}}\right\|^2}_{\text{Global consistency}}+\overbrace{\mu}^{\text{Parameter}}\underbrace{\sum_{i=1}^{n}\left\|f(d_i,q)-f^0(d_i,q)\right\|^2}_{\text{Fit initial scores}}$$

Hongbo Deng, Michael R. Lyu and Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong

Graph-based re-ranking model

- Optimization problem

$$F^* = arg \min_{F \in \mathbb{R}^{+n}} R(F, q, G)$$

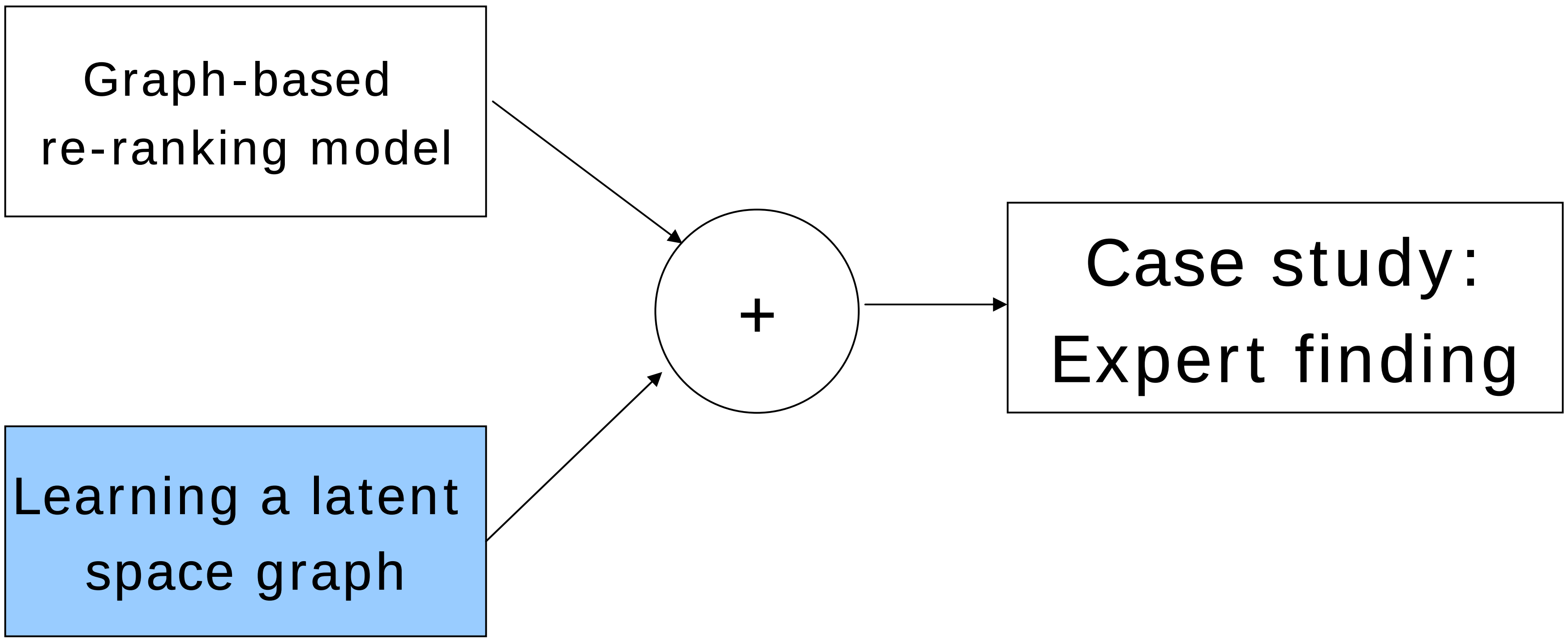
- A closed-form solution

$$F^* = \mu_\beta (I - \mu_\alpha S)^{-1} F^0,$$
$$S = D^{-\frac{1}{2}} W D^{-\frac{1}{2}},$$
$$\mu_\alpha = \frac{1}{1+\mu}, \text{ and } \mu_\beta = \frac{\mu}{1+\mu},$$

$$\boxed{D_{ii} = \sum_j w_{ij}}$$

- Connection with other methods
  - $\mu_\alpha \rightarrow 0$, return the initial scores
  - $\mu_\alpha \rightarrow 1$, a variation of PageRank-based model
  - $\mu_\alpha \in (0, 1)$, combine both information simultaneously

Methodology



Learning a latent space graph

- Objective: incorporate the content with link information (or relational data) simultaneously
  - Latent Semantic Analysis
  - Joint factorization
    - Combine the content with relational data
  - Build latent space graph
    - Calculate the weight matrix W

Latent Semantic Analysis

- Map documents to vector space of reduced dimensionality
- SVD is performed on the matrix $C = U\Sigma V^T$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{bmatrix}$$

- The largest *k* singular values

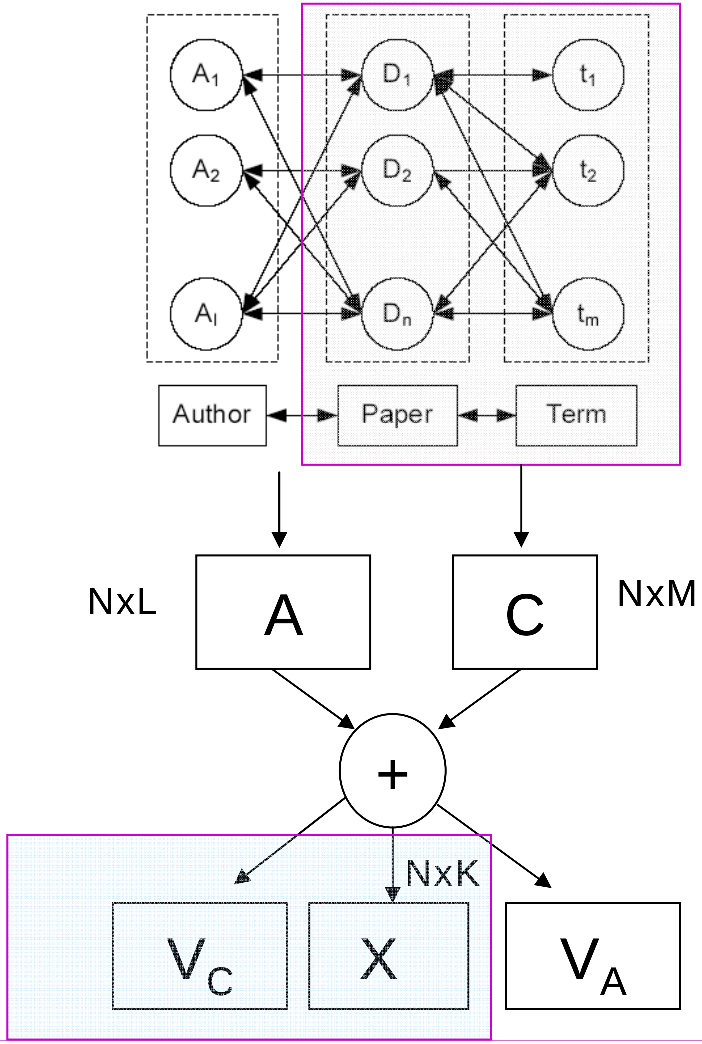
$$\hat{C} = U\hat{\Sigma}V^T \approx U\Sigma V^T = C$$

$$X = U\hat{\Sigma}$$

- Reformulated as an optimization problem

$$\min_{X,V} \| C - XV^T \|_F^2 + \gamma \| V \|_F^2$$

Embedding multiple relational data



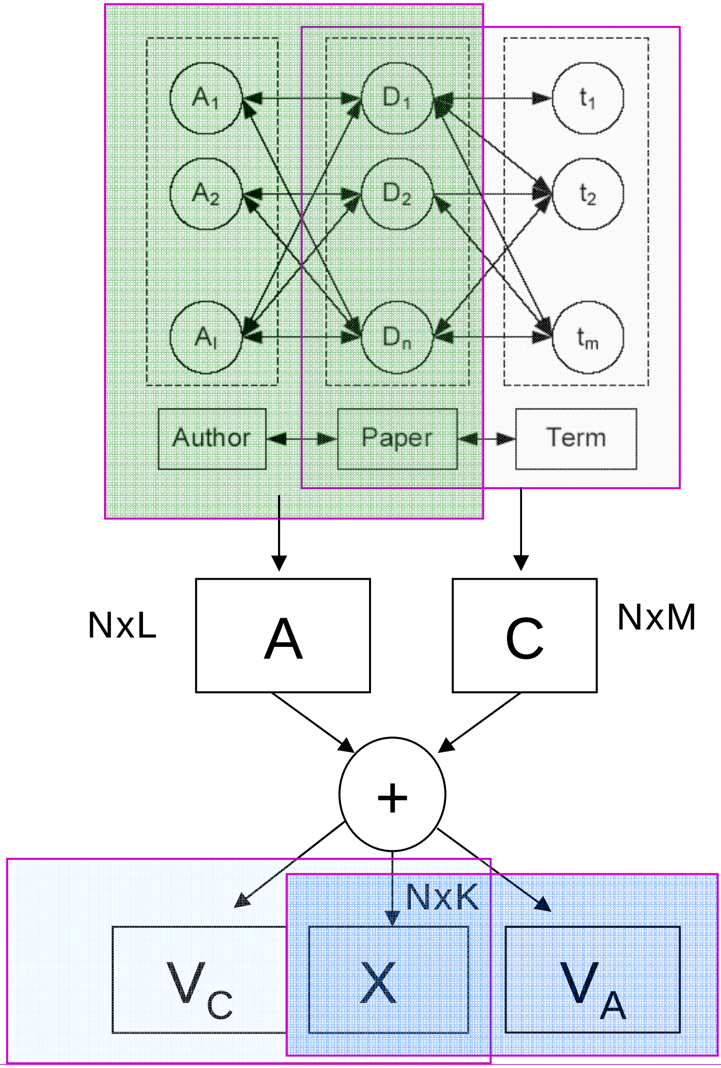
- Taking the papers as an example
  - Paper-term matrix C
  - Paper-author matrix A
- A unified optimization problem

$$
\begin{aligned}
J(X, V_C, V_A) &= \| C - X V_C^T \|_F^2 + \gamma \| V_C \|_F^2 \\
&+ \lambda \left(\| A - X V_A^T \|_F^2 + \gamma \| V_A \|_F^2 \right)
\end{aligned}
$$

Conjugate Gradient

$$
\begin{aligned}
\frac{\partial J}{\partial V_C} &= \left(V_C X^T X - C^T X \right) + \gamma V_C, \\
\frac{\partial J}{\partial V_A} &= \lambda \left(V_A X^T X - A^T X + \gamma V_C \right), \\
\frac{\partial J}{\partial X} &= \left(X V_C^T V_C - C V_C \right) + \lambda \left(X V_A^T V_A - A V_A \right).
\end{aligned}
$$

Embedding multiple relational data



- Taking the papers as an example
  - Paper-term matrix C
  - Paper-author matrix A
- A unified optimization problem

$$J(X, V_C, V_A) = \| C - XV_C^T \|_F^2 + \gamma \| V_C \|_F^2 + \lambda \left(\| A - XV_A^T \|_F^2 + \gamma \| V_A \|_F^2 \right)$$

Conjugate Gradient

$$\frac{\partial J}{\partial V_C} = \left(V_C X^T X - C^T X \right) + \gamma V_C,$$

$$\frac{\partial J}{\partial V_A} = \lambda \left(V_A X^T X - A^T X + \gamma V_C \right),$$

$$\frac{\partial J}{\partial X} = \left(X V_C^T V_C - C V_C \right) + \lambda \left(X V_A^T V_A - A V_A \right).$$

Build latent space graph

- The edge weight w_{ij} is defined

$$w_{ij} = \exp^{-\|x_i - x_j\|^2 / 2\sigma^2}$$

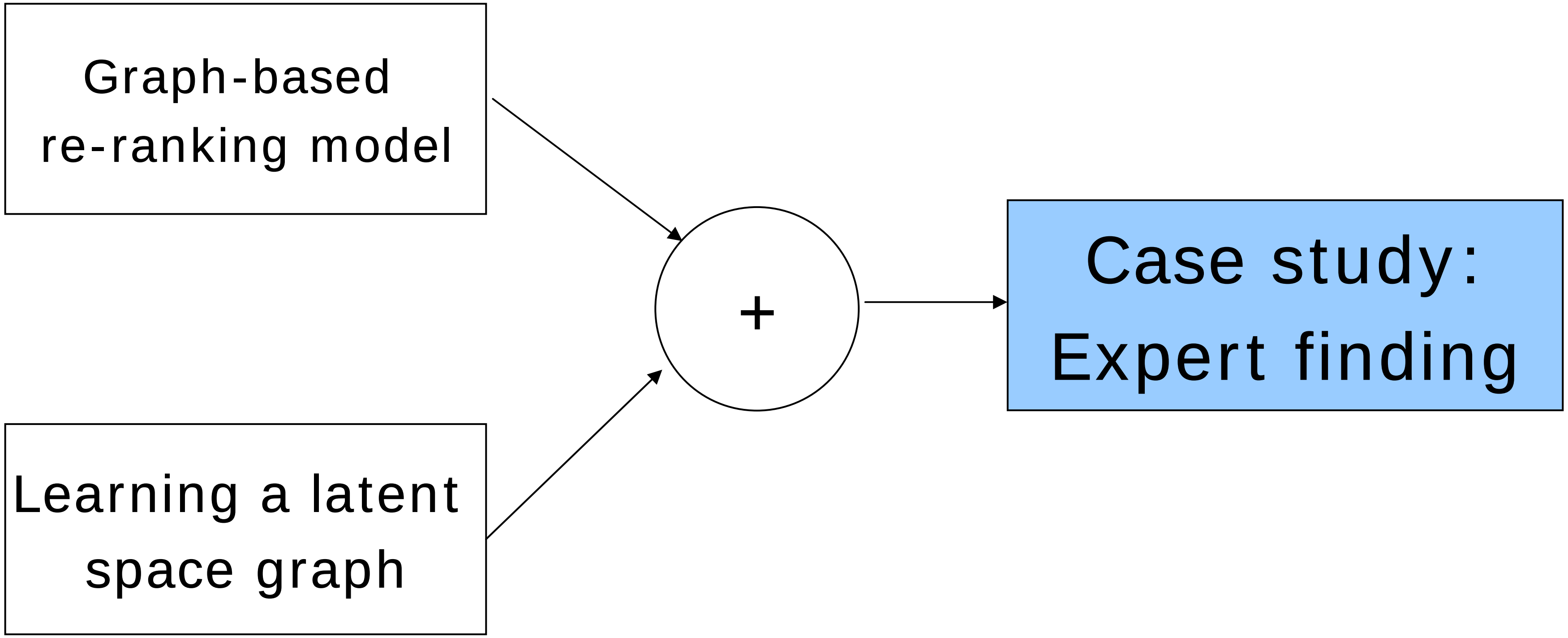
$$\downarrow$$

W

$$\downarrow$$

$$S = D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$$

Methodology



Case study: Application to expert finding

- Utilize statistical language model to calculate the initial ranking scores
  - The probability of a query given a document

$$f^0(d,q) = p(q|\theta_d) = \prod_{t \in q} p(t|\theta_d)^{n(t,q)}$$

- Infer a document model θ_d for each document
- The probability of the query generated by the document model θ_d
- The product of terms generated by the document model (Assumption: each term are independent)

Case study: Application to expert finding

- Expert Finding:
  - Identify a list of experts in the academic field for a given query topic (e.g., "data mining" → "Jiawei Han, etc")
  - Publications as representative of their expertise
  - Use DBLP dataset to obtain the publications

- Authors have expertise in the topic of their papers
  - Overall aggregation of their publications

$$P(ca, q) = \sum_{d \in \mathbb{D}_{ca}} \frac{1}{n_d} f(d, q)$$

  - Refine the ranking scores of papers, then aggregate the refined scores to re-rank the experts

Case study: Application to expert finding

Online: Expert finding application
Input: Given a query q,
Perform:

1. Calculate the initial ranking scores F^0 based on the language model
$f^0(d,q) = \prod_{t \in q} p(t|\theta_d)^{n(t,q)}$;

2. Extract the top-ranked documents as a subset $\hat{\mathbb{D}}$, the corresponding ranking scores $\hat{F}^0$ and the subgraph $\hat{S}$;

3. Re-rank with the subgraph to approximate
$\hat{F}^* = (I - \mu_\alpha \hat{S})^{-1} \hat{F}^0$;

4. Aggregate the expertise $P(ca, q)$ for authors.

Output: Return the ranked experts $\{ca_1, ca_2, ..., ca_k\}$

Experiments

DBLP Collection

- A subset of the DBLP records (15-CONF)
- Statistics of the 15-CONF collection

Property	#of entities
paper	31437
author	30901
term	16938

Benchmark Dataset

- A benchmark dataset with 16 topics and expert lists

Query	#Expert
boosting	47
information retrieval	17
information extraction	20
intelligent agents	27
machine learning	42
ontology alignment	33
planning	30
privacy preservation	18
reinforcement learning	16
semantic web	44
sensor RFID data management	13
skyline	12
stream	16
support vector machine	22
semi-supervised learning	21
kernel method	22

Evaluation Metrics

- Precision at rank n (P@n):

$$P@n = \frac{\#\ \text{relevant candidates in top } n \text{ results}}{n}$$

- Mean Average Precision (MAP):

$$AP = \frac{\sum_{n=1}^{N}(P@n * \mathrm{rel}(n))}{R}$$

- Bpref: The score function of the number of non-relevant candidates

$$\mathrm{bpref} = \frac{1}{R}\sum_{r=1}^{N}(1 - \frac{\#n \text{ ranked higher than } r}{R})$$

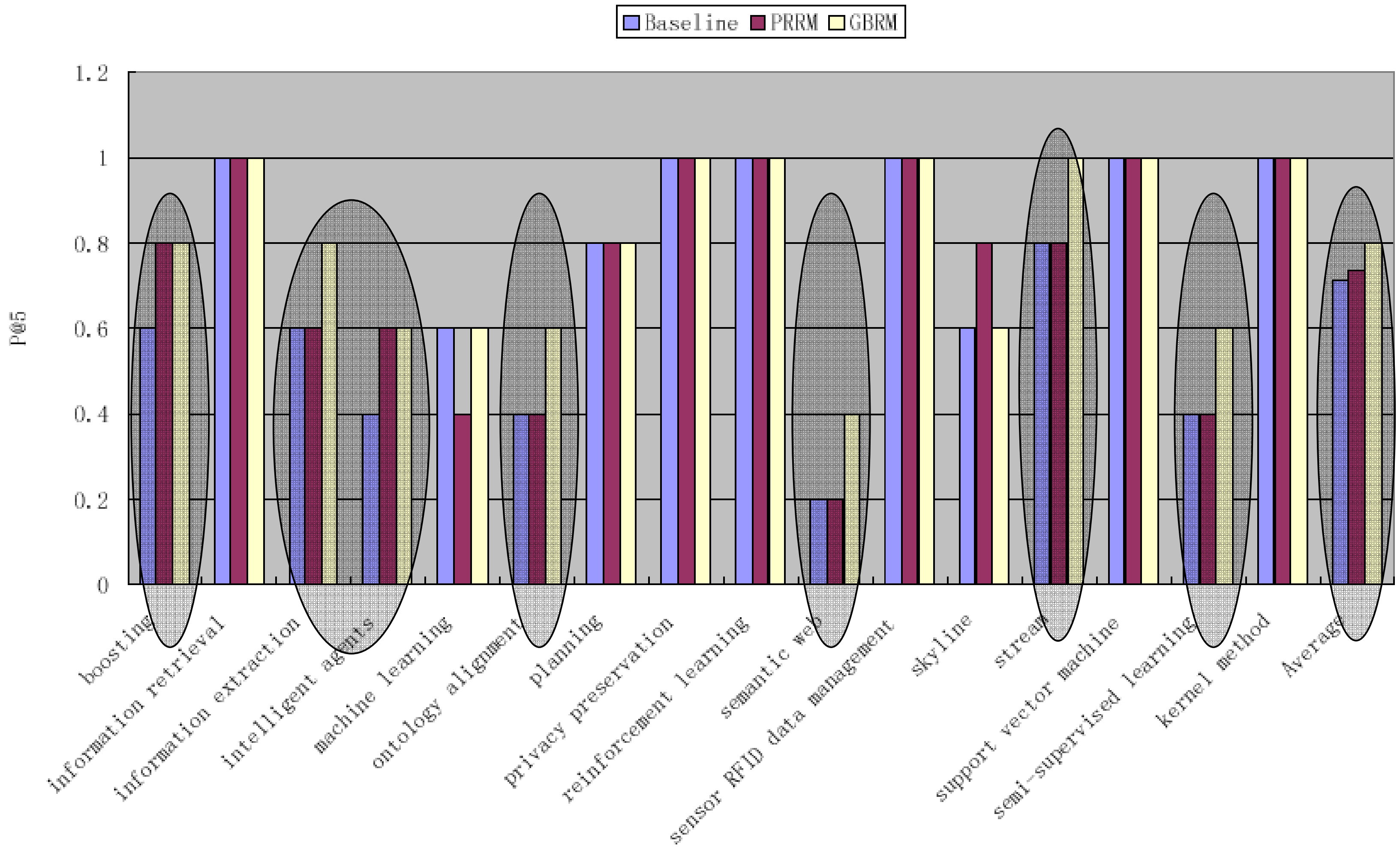
Preliminary Experiments

- Evaluation results (%)

	P@5	P@10	P@20	MAP	bpref
Baseline	71.25	68.12	55.62	43.72	48.61
PRRM	73.75	65.63	51.56	39.94	44.85
GBRM	**80.00**	**73.13**	**56.87**	**45.83**	**50.70**

  - PRRM may not improve the performance
  - GBRM achieve the best results

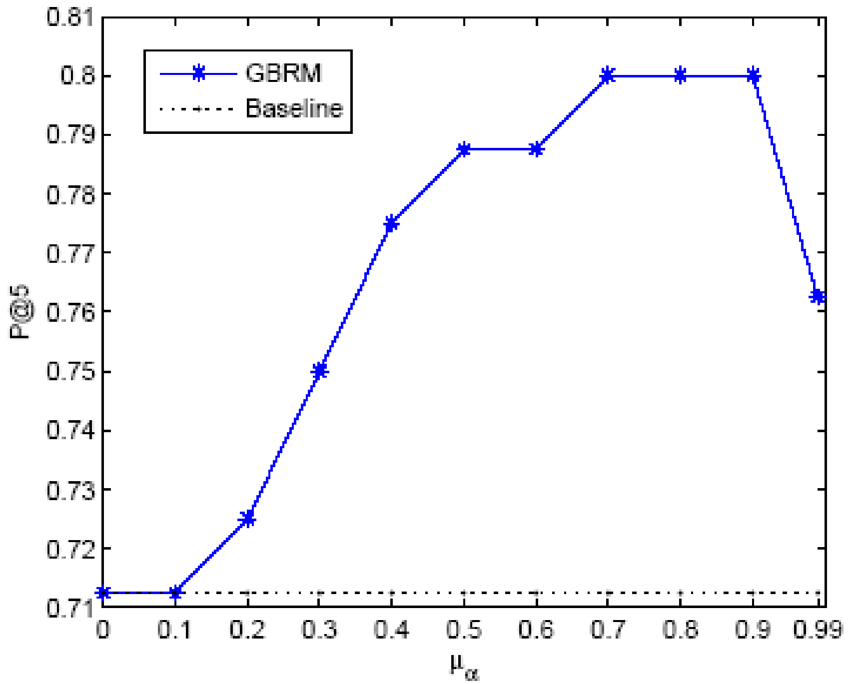
Details of the results



Hongbo Deng, Michael R. Lyu and Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong

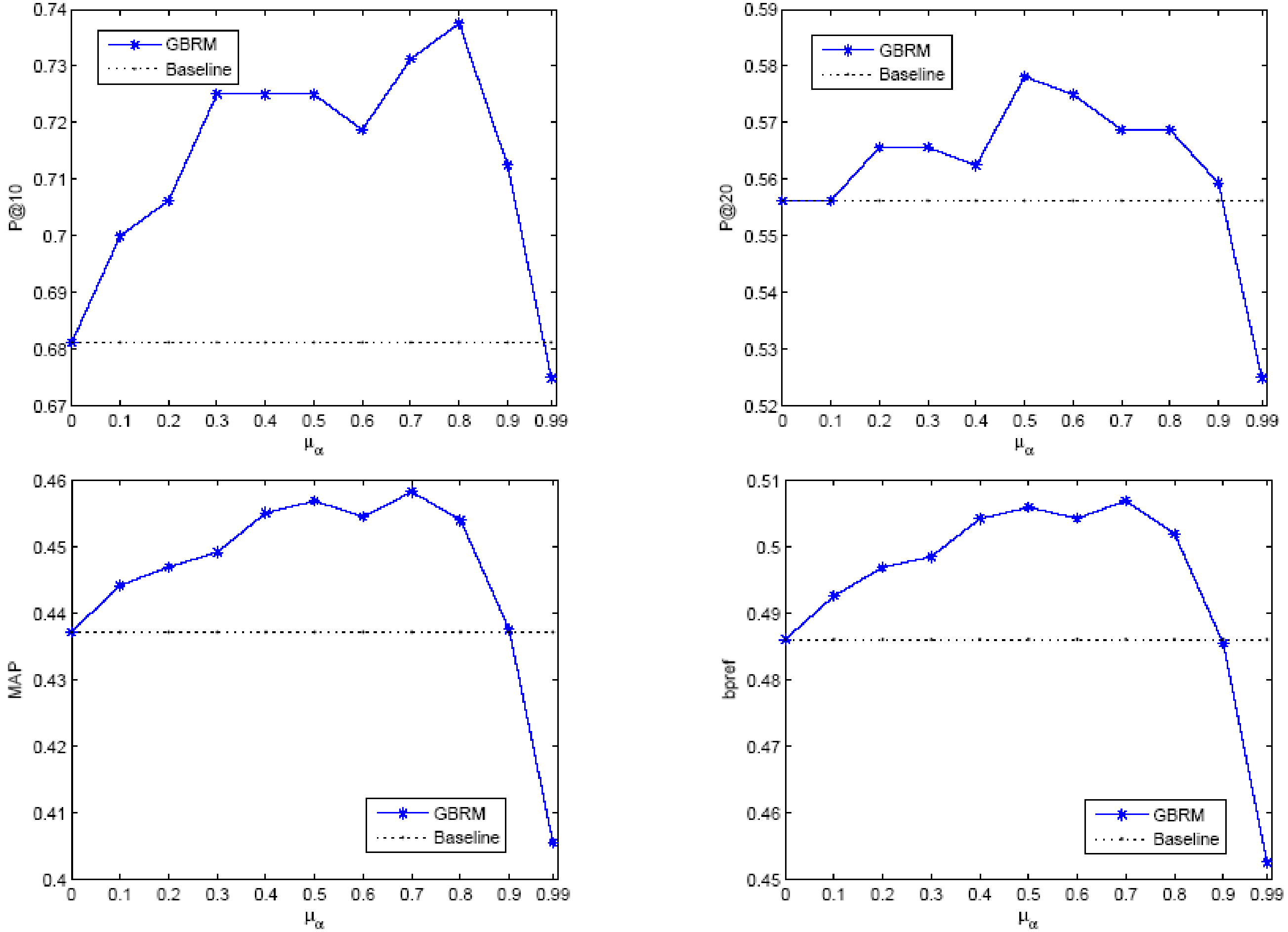
Effect of parameter μ_α

- $\mu_\alpha \rightarrow 0$, return the initial scores (baseline)
- $\mu_\alpha \rightarrow 1$, discard the initial scores, consider the global consistency over the graph



Hongbo Deng, Michael R. Lyu and Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong

Effect of parameter μ_α



Robust, achieve the best results when $\mu_a \rightarrow (0.5, 0.7)$

Effect of graph construction

- Different dimensionality (k_d) of the latent feature, which is used to calculate the weight matrix W

	P@5	P@10	P@20	MAP	bpref
TFIDF	77.50	70.63	57.50	44.70	49.52
$k_d = 20$	72.50	71.88	**58.13**	45.36	50.44
$k_d = 50$	77.50	71.88	57.50	45.77	**50.86**
$k_d = 100$	**80.00**	**73.13**	56.88	**45.84**	50.70

  - Become better for greater k_d, *because* higher dimensional space can better capture the similarities
  - $k_{d,}$ = 50 → achieve better results than *tf.idf*

Effect of graph construction

- Different number of nearest neighbors (k_{nn})

	P@5	P@10	P@20	MAP	bpref	Time
$k_{nn} = 5$	78.75	71.88	56.88	45.03	49.85	0.98s
$k_{nn} = 10$	80.00	73.13	56.88	45.84	50.70	2.11s
$k_{nn} = 20$	80.00	71.88	57.19	45.80	50.60	5.53s
$k_{nn} = 30$	78.75	71.88	56.88	45.88	50.65	9.06s
$k_{nn} = 40$	78.75	71.88	56.56	45.33	50.13	13.21s

  - Tends to degrade a little with increasing k_{nn}
  - $k_{nn} = 10 \rightarrow$ achieve the best results
  - Average processing time: increase linearly with the increase of k_{nn}

Conclusions and Future Work

- Conclusions
  - Leverage the graph-based model for the query-dependent ranking problem
  - Integrate the latent space with the graph-based re-ranking model
  - Address expert finding task in a the academic field using the proposed method
  - The improvement in our proposed model is promising
- Future work
  - Extend our framework to consider more features
  - Apply the framework to other applications and large-scale dataset

Q&A

Thanks!