WSDM 2015 workshop on
Vertical Search Relevance
Feb. 6, 2015
Shanghai, China

Post Processing of Ranking in Search

Jun Xu

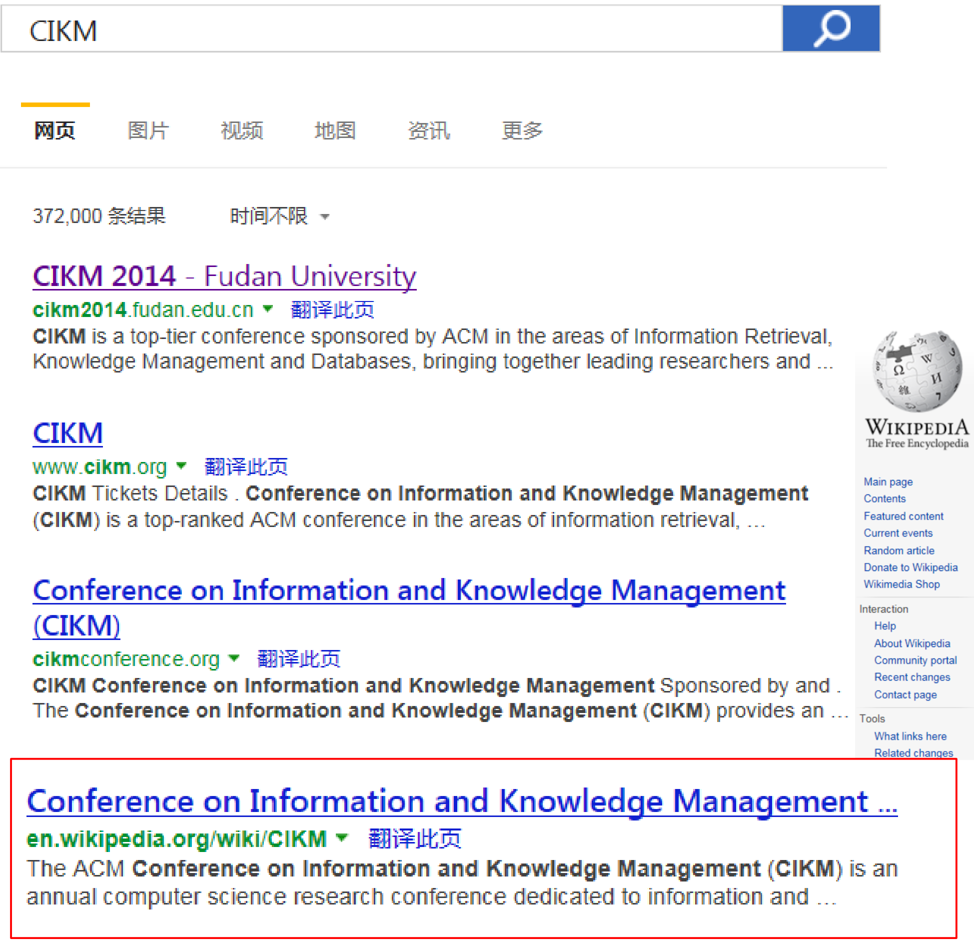
Institute of Computing Technology,
Chinese Academy of Sciences

Joint work with Fangzhao Wu, Hang Li, and Xin Jiang

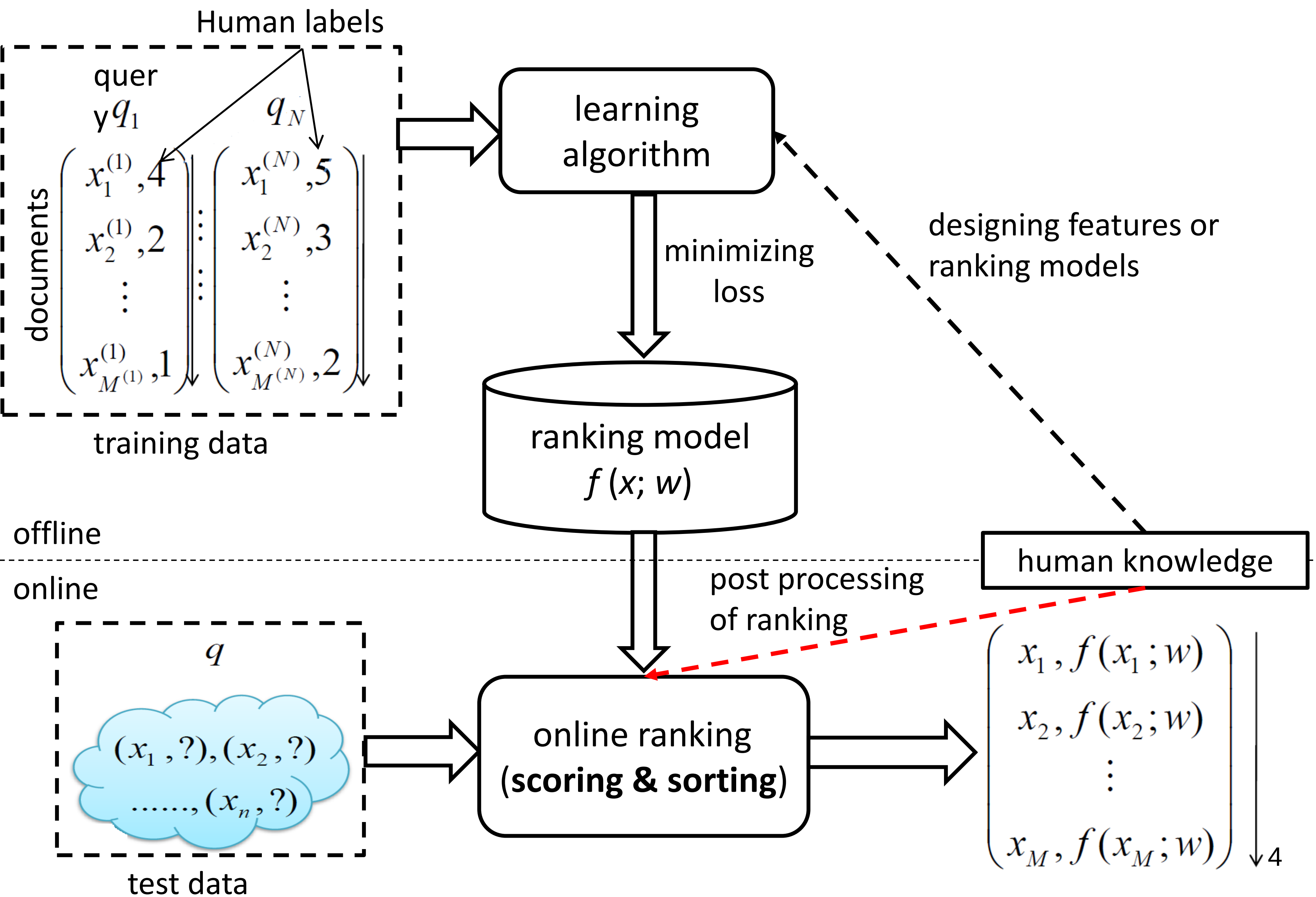
Outline

- Post processing of ranking
- Ranking optimization with constraints
- Summary

One may want to 'twist' relevance ranking



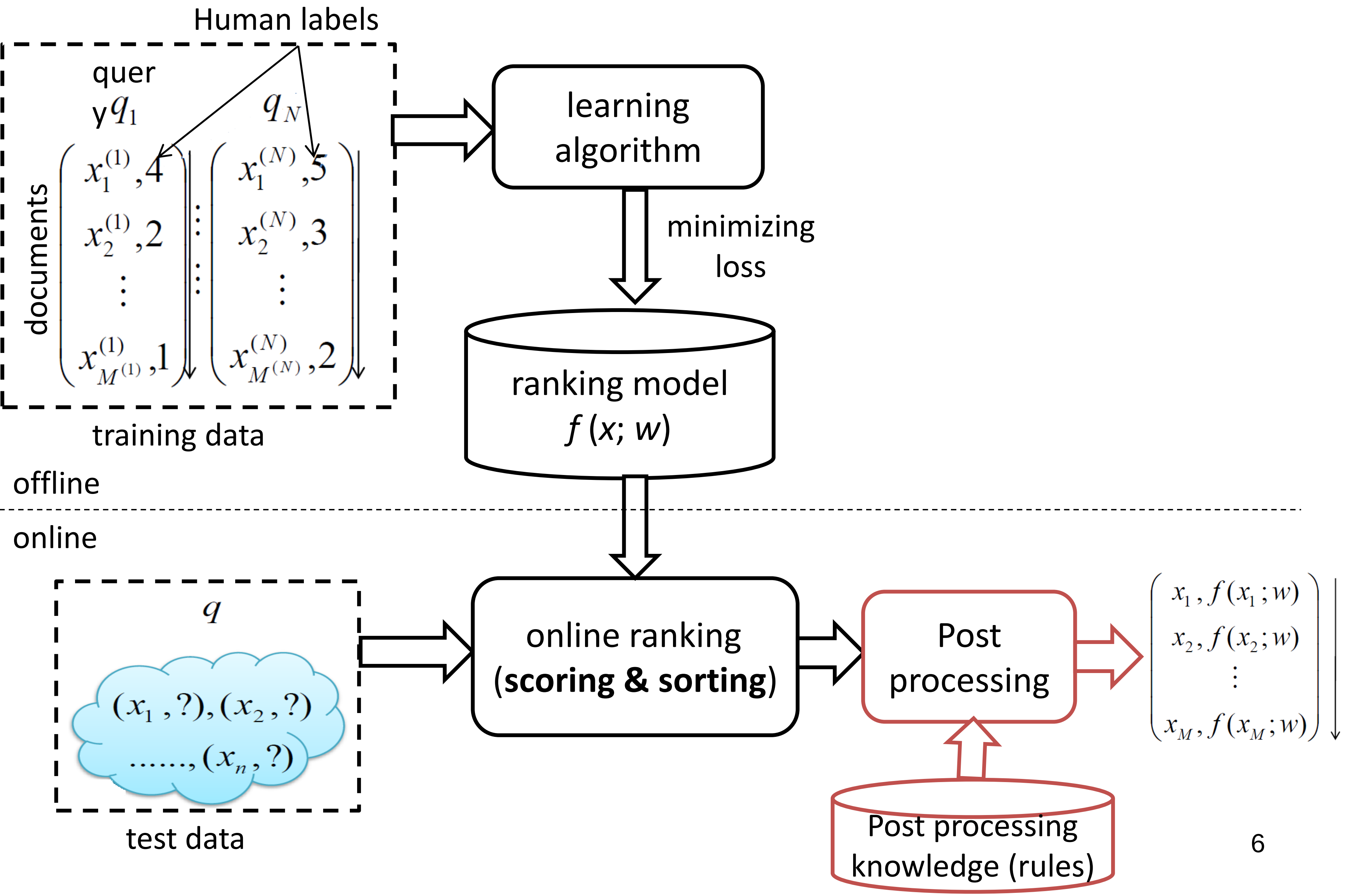
Learning Approach to Ranking



Incorporating Human Knowledge

- Designing features or ranking models
  - Indirect
  - Limited to cross-query knowledge, for generalizing to other queries; It is difficult, costly, or even impossible to implement in features and models to incorporate some types of knowledge
  - Modify both offline and online components
- Post processing of ranking
  - Direct (apply on test queries directly)
  - Can incorporate query/user/context dependent knowledge
  - Only modify online component

Post Processing of Ranking



Post Processing Knowledge (Rules)

- Query dependent
  - Specific query type: if the query is a name, promote the corresponding personal homepage
  - Specific query: if the query is “Microsoft”, promote http://www.microsoft.com/ to rank 1
- User dependent (personalization)
  - Query: “Michael Jordan”
    - For basketball fan: promote the Wikipedia entry of the basketball player
    - For CS researcher: promote the Wikipedia entry of the professor at the UC Berkeley

Post Processing Knowledge (cont')

- Context dependent (session-based)

	Query 1: "homes for rent in atlanta"		Query 2: "houses for rent in atlanta"
×	**Atlanta homes for rent - home rentals - houses for ren...** Rentlist is directory of Atlanta home rentals featuring links to... *http://www.rentlist.net*		**Atlanta homes for rent - home rentals - houses for ren...** Rentlist is directory of Atlanta home rentals featuring links to... *http://www.rentlist.net*
	Homes For Rent, lease in Atlanta suburbs. Can't sell ... Atlanta homes for rent, homes for lease in Gwinnett and north... *http://atlantahomesforrent.com*		**Homes for Rent in Atlanta, GA** Houses, Apartments and Homes for Rent in Atlanta, GA Find ... *http://www.usrentallistings.com/ga/atlanta*
	Rentals.com - Homes for Rent, Apartments, Houses ... Atlanta Home Rentals; Austin Home Rentals; Charlotte Home... *http://www.rentals.com*		**Atlanta Home Rentals, Homes for Rent in Atlanta ...** Atlanta Rentals - Homes for Rent in Atlanta, Apartments, Re... *http://www.rentals.com/Georgia/Atlanta*
×	**Atlanta Home Rentals, Homes for Rent in Atlanta ...** Atlanta Rentals - Homes for Rent in Atlanta, Apartments, Re... *http://www.rentals.com/Georgia/Atlanta*		**Homes For Rent, lease in Atlanta suburbs. Can't sell ...** Atlanta homes for rent, homes for lease in Gwinnett and north... *http://atlantahomesforrent.com*
	Homes for Rent in Atlanta, GA Houses, Apartments and Homes for Rent in Atlanta, GA Find ... *http://www.usrentallistings.com/ga/atlanta*	×	**Atlanta Homes for Rent, Rental Properties, Houses for ...** Search for Homes for Rent in Atlanta, Georgia for free. View li... *www.rentalhouses.com/find/GA/AtlantaArea/ATLANTA*

From Xiang et al., SIGIR' 10

Post Processing Knowledge (cont')

- Document (website) dependent
  - Example rule: if webpage from one site is ranked at top, webpages from the other site will be demoted

编辑 | 讨论 | 分享 | 1832

中科院计算技术研究所 认领 HOT

中国科学院计算技术研究所（简称计算所）创建于1956年，是中国第一个专门从事计算机科学技术综合性研究的学术机构。计算所研制成功了我国第一台通用数字电子计算机，并形成了我国高性能计算机的研发基地，我国首枚通用CPU芯片也诞生在这里。

编辑摘要

目录
1 简介
2 所长致辞
3 现任领导
4 发展战略
5 大事记

中科院计算技术研究所 - 简介

中国科学院计算技术研究所（简称计算所）创建于1956年，是中国第一个专门从事计算机科学技术综合性研究的学术机构。计算所研制成功了我国第一台通用数字电子计算机，并形成了我国高性能计算机的研发基地，我国首枚通用CPU芯片也诞生在这里。

计算所是我国计算机事业的摇篮。随着学科与技术发展，从计算所陆续分离出中科院微电子学研究所、计算中心、软件所和网络中心等多个研究机构，及联想、曙光等高技术企业。

计算所是中国科学院"知识创新工程"首批试点单位。正在发展成为由本部核心所和建在上海、苏州、宁波、东莞等地的若干个分部组成的网络型研究所。

计算所本部现有三个研究部和两个研究中心。系统结构研究部主要从事与计算机系统相关的研究，如计算机体系结

http://www.baike.com/wiki/中科院计算技术研究所

中国科学院计算技术研究所 编辑

中科院计算所一般指中国科学院计算技术研究所

本词条缺少信息栏，补充相关内容使词条更完整，还能快速升级，赶紧来编辑吧！

中国科学院计算技术研究所(Institute of Computing Technology, Chinese Academy of Science,简称ICT)创建于1956年，坐落于北京海淀区，中关村南路科学院六号。是中国第一个专门从事计算机科学技术综合性研究的学术机构。计算所研制成功了中国第一台通用数字电子计算机，并形成了中国高性能计算机的研发基地，中国首枚通用CPU芯片也诞生在这里。

目录
1 所况简介
2 现任领导
· 孙凝晖
· 李锦涛
· 隋雪青
· 陈熙霖
· 张光辉
3 学术委员会
4 研究实体
· 计算机系统研究部
· 网络科学与技术研究部
· 智能信息处理研究部
5 管理部门
· 综合处
· 人力资源处
· 科研处
· 技术发展处
· 研究生部
· 财务资产处
6 支撑部门
· 科研支撑中心
· 期刊编辑部
· 图书馆
7 计算所分部
· 上海分部
· 苏州分部
· 烟台分所
8 光辉历程
9 科研概况
10 成果介绍
11 科技奖励
12 质量认证
13 研究生教育

所况简介 编辑

计算所是中国计算机事业的摇篮。随着学科与技术发展，从计算所陆续分离出中科院微电子研究所、计算中心、软件研究所

http://baike.baidu.com/view/730187.htm

Heuristic Approaches

- Widely used in real search systems, however
  - Rules may be ambiguous, e.g., the document should be ranked at top three positions, no specific position is decided
  - Rules might be contradictory, e.g., two rules want to rank different documents to top 1
  - Different orders of applications of rules might yield different ranking results. The later one has higher priority
  - Hard to balance between application of rules and preservation of the original ranking list
  - Hard to manage the old/new rules
- Difficult to formalize in a theoretically sound, effective, and efficient way

Outline

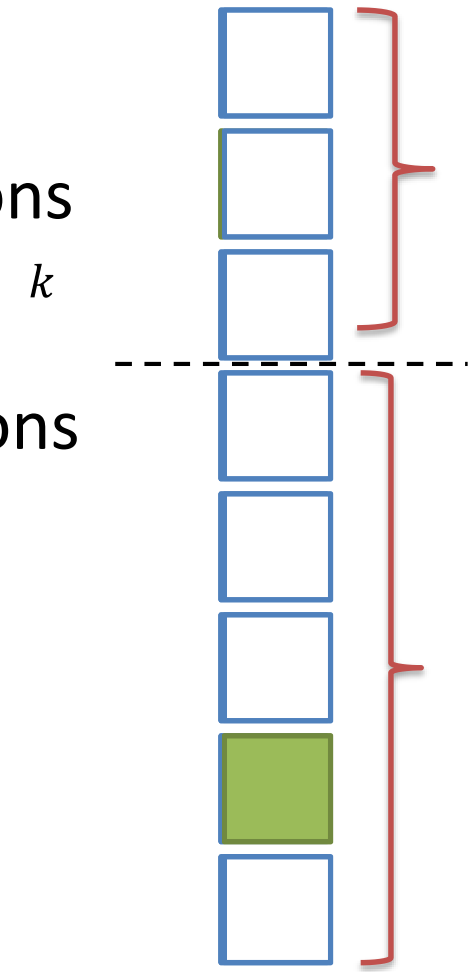
- Motivation of post processing of ranking
- Ranking optimization with constraints (Wu et al., CIKM 2014)
- Conclusion

Main Idea

- Traditional approaches
  - Mainly based on heuristic rules
  - No principled approach
- Our work
  - Formalizes as a constrained optimization problem
    - Constraints: post-processing rules
    - Object function: tradeoff between original ranking and rules
  - Implementation with Bradley-Terry model

Covered Constraints

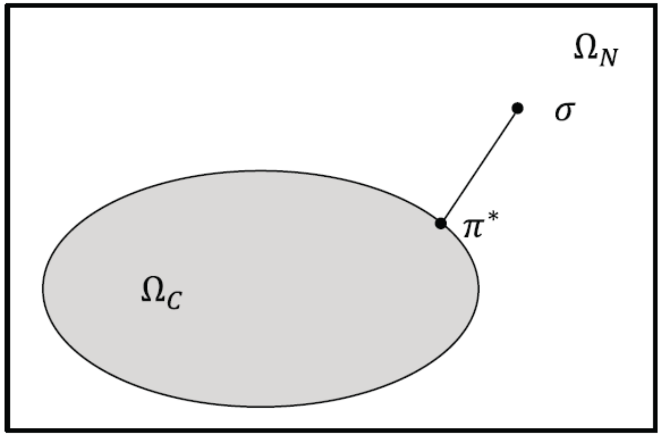
- Top-k constraint
  - A document should be at top k positions
- Not-top-k constraint
  - A document cannot be at top k positions



Related Work

- Post ranking with heuristics
  - **Result diversification**: re-ranking after a ranking list based on relevance is created [Dou et al., ‘11; Vee et al., ‘08]
  - **Personalized search**: client side re-ranking based on user interest [Radlinski & Dumais, ‘ 06; Sugiyama et al., ‘04; Teevan et al., ‘05]
  - **Context aware ranking**: demoting clicked URL in current search result, if it was clicked in the previous search in the same session [Xiang et al., ‘10]
- Probabilistic models for ranking
  - Plackett-Luce model: stage-wise generative model [Luce ‘75]
  - Mallows model: distance based [Mallows ‘57]
  - Bradely-Terry model: pairwise comparisons [Bradley & Terry, ‘52]

Ranking Optimization with Constraints



$$\min_{\pi \in \Omega_N} L(\sigma, \pi) + \lambda \cdot R(\mathcal{C}, \pi)$$

difference between σ and π

violation of constraints

- **Constraints:** rules for post ranking. $\mathcal{C} = \{c_i(\cdot)\}$, $c_i\colon \Omega_N \to \{0, 1\}$
- **Objective function:** trade-off between adherence to the *original ranking list* and satisfaction of the *constraints*

Probabilistic Approach

- Introducing probabilistic ranking model M and $\pi = \text{argmax}_{\tau} P(\tau|M)$

$$\min_M L(\sigma, M) + \lambda \cdot R(\mathcal{C}, M)$$

- Define
  - $L(\sigma, M) = -\log P(\sigma|M)$
  - $R(\mathcal{C}, M) = -\log P(\mathcal{C}|M)$
- Two steps
  - Estimating M

$$\min_M -\log P(\sigma|M) - \lambda \cdot \log P(\mathcal{C}|M)$$

  - Getting optimal ranking list

$$\pi^* = \text{argmax}_{\pi \in \Omega_N} P(\pi|M)$$

Using Bradley-Terry Model

- Represents distribution of permutation by making pairwise comparisons

$$p_{ij} = P\{(i,j)\} = \frac{\theta_i}{\theta_i + \theta_j}$$

- Probability of a permutation

$$P(\sigma|M) \propto \prod_{(i,j):\sigma(i)<\sigma(j)} p_{ij} = \prod_{(i,j):\sigma(i)<\sigma(j)} \frac{\theta_i}{\theta_i + \theta_j}$$

- Probability of a constraint set

$$P(\mathcal{C}|M) \propto \prod_{c\in\mathcal{C}} \prod_{(i,j)\in P^c} p_{ij} = \prod_{c\in\mathcal{C}} \prod_{(i,j)\in P^c} \frac{\theta_i}{\theta_i + \theta_j}$$

P^c: set of preference pairs derived from constraint c
top-k constraint: $P^c = \{(i,j)|j: \sigma(j) > k\}$
not-top-k constraint: $P^c = \{(j,i)|j: \sigma(j) \leq k\}$

Objective Function

$$\min_{M} -\log P(\sigma|M) - \lambda \cdot \log P(\mathcal{C}|M)$$

$$P(\sigma|M) \propto \prod_{(i,j):\sigma(i)<\sigma(j)} \frac{\theta_i}{\theta_i+\theta_j},$$

$$P(\mathcal{C}|M) \propto \prod_{c\in\mathcal{C}} \prod_{(i,j)\in P^c} \frac{\theta_i}{\theta_i+\theta_j}$$

$$\min_{\Theta} f(\Theta) = -\sum_{(i,j):\sigma(i)<\sigma(j)} \log \frac{\theta_i}{\theta_i+\theta_j} - \sum_{c\in\mathcal{C}} \left(\rho^c \cdot \sum_{(i,j)\in\mathcal{P}^c} \log \frac{\theta_i}{\theta_i+\theta_j} \right)$$

$$\text{subject to } \forall i : \theta_i > 0, \sum_{i=1}^{N} \theta_i = 1,$$

$$\theta_i = \exp\{s_i\}, s_i \in R$$

$$\min_{\mathcal{S}} f(\mathcal{S}) = \sum_{(i,j):\sigma(i)<\sigma(j)} (\log(e^{s_i} + e^{s_j}) - s_i) + \sum_{c\in\mathcal{C}} \left(\rho^c \cdot \sum_{(i,j)\in\mathcal{P}^c} (\log(e^{s_i} + e^{s_j}) - s_i) \right)$$

THEOREM 4.1. *$f(\mathcal{S})$ is a convex function.*

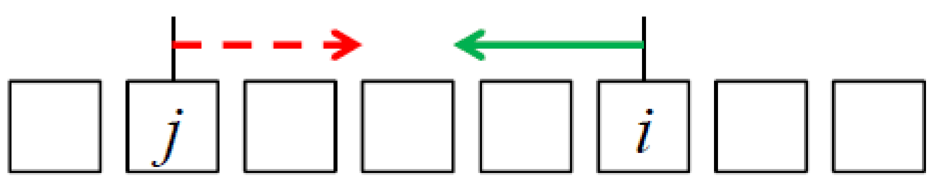
Optimizing with Gradient Descent

$$\frac{df}{ds_i} = \left(\sum_{j:\sigma(j)<\sigma(i)} \frac{e^{s_i}}{e^{s_i}+e^{s_j}} - \sum_{j:\sigma(i)<\sigma(j)} \frac{e^{s_j}}{e^{s_i}+e^{s_j}} \right) + \sum_{c\in\mathcal{C}} \rho^c \left(\sum_{j:(j,i)\in\mathcal{P}^c} \frac{e^{s_i}}{e^{s_i}+e^{s_j}} - \sum_{j:(i,j)\in\mathcal{P}^c} \frac{e^{s_j}}{e^{s_i}+e^{s_j}} \right)$$

Demote i Promote i

$$\mathcal{S}^{(t)} = \mathcal{S}^{(t-1)} - \gamma^{(t)} \left. \frac{\partial f}{\partial \mathcal{S}} \right|_{\mathcal{S}=\mathcal{S}_{t-1}}$$

- Intuitive explanation: given a pair (i, j), i be pushed upward and j be pushed downward with identical force strengths



Ranking Optimization Algorithm

Algorithm 1 Ranking Optimization Algorithm

Require: Initial ranking σ, constraints $\mathcal{C}$, and shrinkage rate $0 < \alpha < 1$

$\mathcal{S}^{(0)} \leftarrow$ random values
$t \leftarrow 1$
repeat
  $\nabla \mathcal{S} = \frac{\partial f}{\partial \mathcal{S}}\big|_{\mathcal{S}=\mathcal{S}^{(t-1)}}$ {Equation (6)}
  $\gamma \leftarrow 1$
  {search optimal step size using backtracking}
  while $f(\mathcal{S}^{(t-1)} - \gamma \nabla \mathcal{S}) > f(\mathcal{S}^{(t-1)}) - \frac{\gamma}{2} \| \nabla \mathcal{S} \|^2$ **do**
    $\gamma \leftarrow \alpha \gamma$
  end while
  $\mathcal{S}^{(t)} \leftarrow \mathcal{S}^{(t-1)} - \gamma \nabla \mathcal{S}$ {Equation (7)}
  $t \leftarrow t + 1$
until convergence
return $\Theta = \{\frac{e^{s_1}}{Z}, \cdots, \frac{e^{s_N}}{Z}\}$, where $Z = \sum_{n=1}^{N} e^{s_n}$

THEOREM 4.2. *Algorithm 1 converges in finite steps and the convergence rate is $O(\frac{1}{\epsilon})$, where $\epsilon > 0$ is the tolerance.*

Experimental Settings

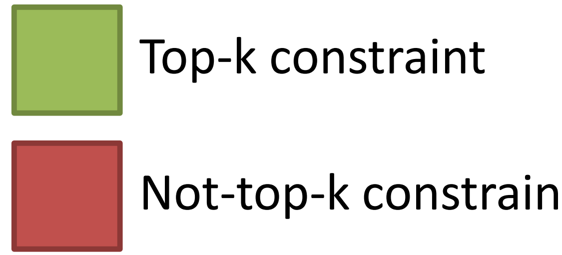
- Datasets

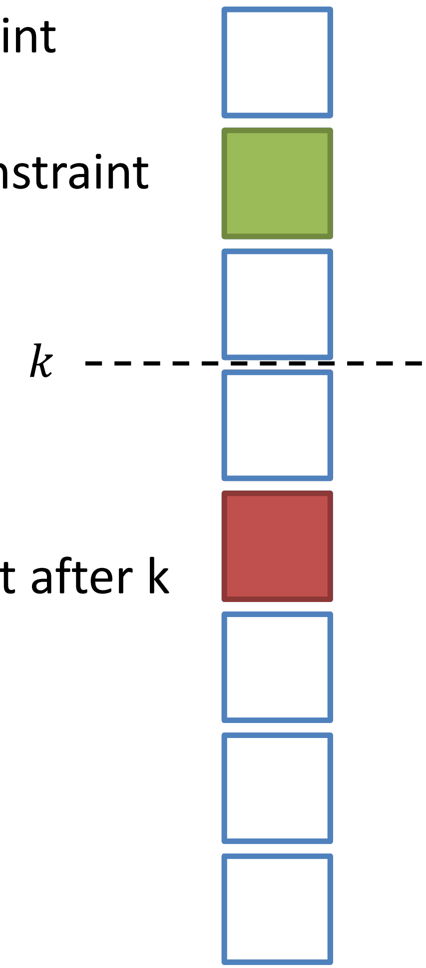
Table 1: Statistics of datasets.

dataset	# queries	#documents	#relevance levels
MQ2007	1692	69623	3
MQ2008	784	15211	3
OHSUMED	106	16140	3
.Gov	50	49058	2
Enterprise	183	5464	3

- Basic ranking model: LambdaMART
- Constraints construction
  - Top-k constraint (k = 1, 3, 5): for each query, sort documents according to labels and randomly select one document from top k positions
  - Not-top-k constraint (k = 5, 10): for each query, sort documents according to labels and randomly select one document from the positions after k positions

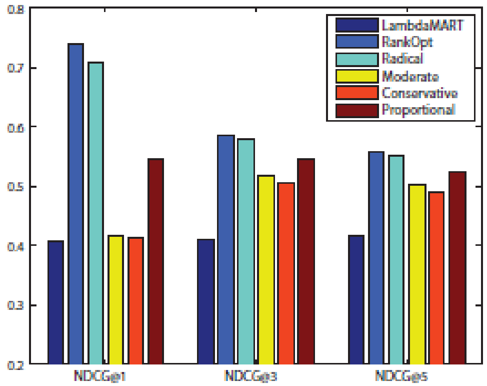
Experimental Settings



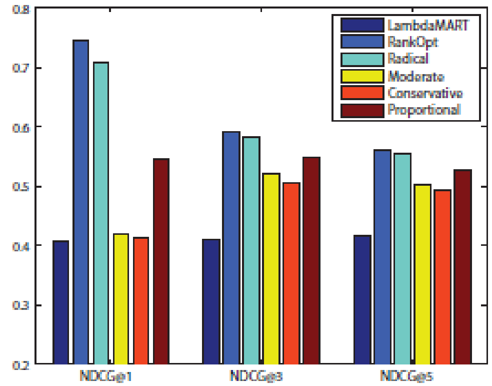


- Baselines
  - **Radical**
    - Top-k constraint → top one position
    - Not-top-k constraint → bottom position
  - **Moderate**
    - Top-k constraint → middle of the top *k* positions
    - Not-top-k constraint → middle of the remaining list after k
  - **Conservative**
    - Top-k constraint → the position of k
    - Not-top-k constraint → the position of k + 1.
  - **Proportional**
    - Top-k constraint → the position of $\left\lceil k \times \frac{pos}{N} \right\rceil$
    - Not-top-k constraint → the position of $\left\lceil k + pos\left(1 - \frac{k}{N}\right) \right\rceil$

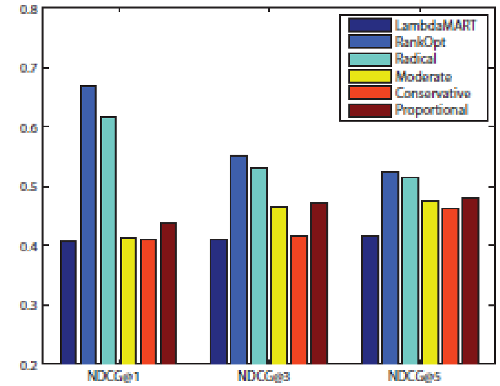
Experimental Results



(a) top-3, not-top-5, ρ^t=100, ρ^n=10

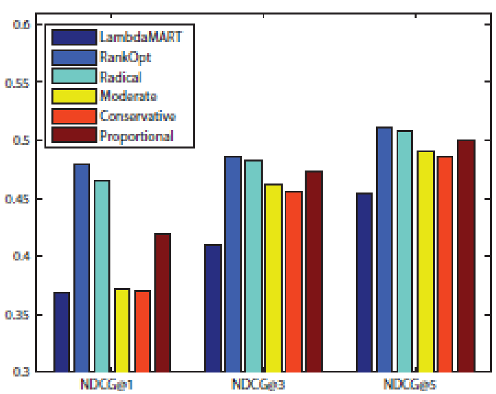


(b) top-3, not-top-10, ρ^t=100, ρ^n=10

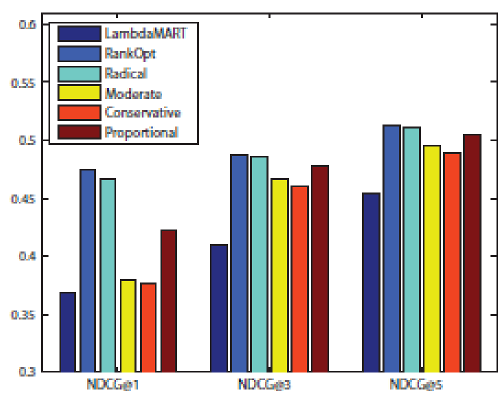


(c) top-5, not-top-10, ρ^t=100, ρ^n=10

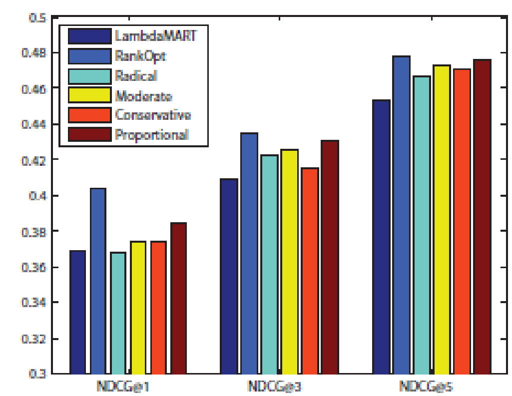
MQ2007



(a) top-3, not-top-5, ρ^t=10, ρ^n=0



(b) top-3, not-top-10, ρ^t=10, ρ^n=10



(c) top-5, not-top-10, ρ^t=10, ρ^n=10

MQ2008

Average Running Time per Query

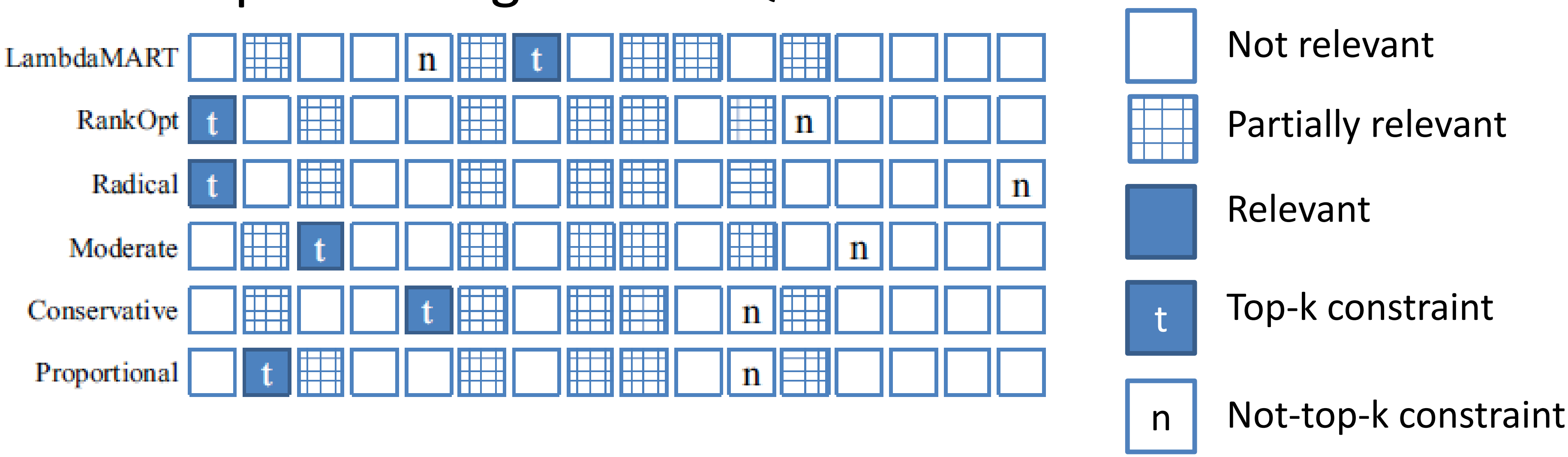
Table 2: Average time (in milliseconds) of ranking optimization in setting of (top-5, not-top-10).

	MQ2008	MQ2007	OHSUMED	.Gov	Enterprise
time	4.24	6.85	134.53	70.06	6.45

- Tested on a Laptop PC with 2.4GHZ CPU and 4GB memory
- For most queries, the algorithm converges within 10 iterations
- Ranking optimization can be performed online

Case Study 1: How Ranking Optimization Works

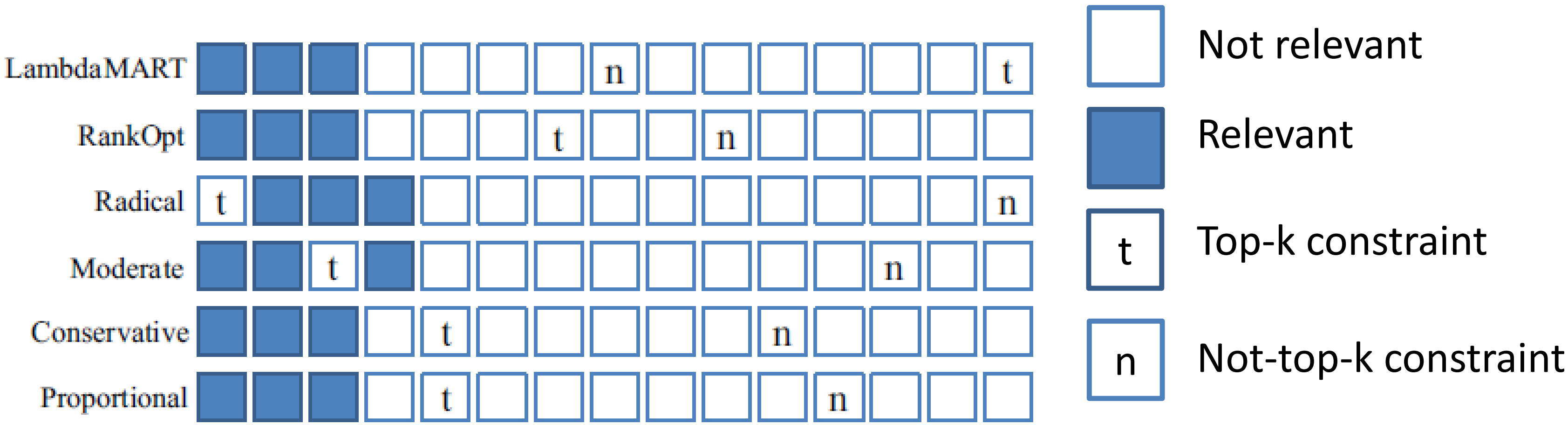
- Example ranking from MQ2008



  - RankOpt promoted the relevant document and demoted the not relevant documents
  - RankOpt outperformed baselines of Moderate, Conservative, and Proportional, when constraints are correct

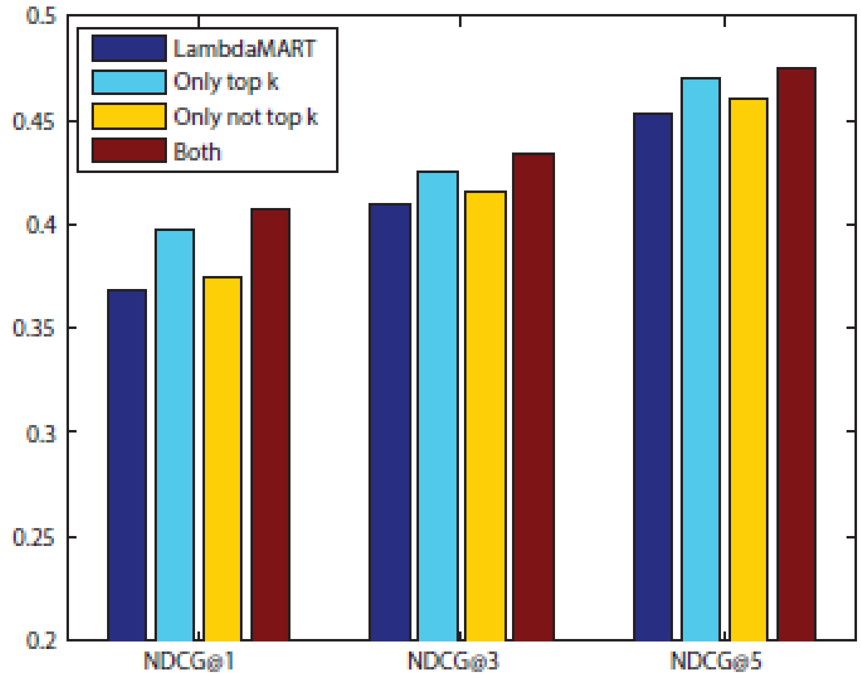
Case Study 2: How Ranking Optimization Works

- Example ranking from MQ2008



  - RankOpt outperformed Radical method, if constraints contain noise
  - RankOpt made good trade-off between constraints and original ranking

Discussion: Constraint Types



- Top-k and not-top-k constraints individually improved the ranking performances
- Performances be further improved when both are used
- RankOpt can leverage multiple types of constraints

Outline

- Motivation of post processing of ranking
- Ranking optimization with constraints
- Summary

Summary

- Post-processing of ranking is important for search
- Heuristic approaches have limitations
- Our preliminary work makes use of Bradley-Terry model for handling the top-k and not-top-k rules
- Next step
  - Defining and incorporating other types of constraints into the framework, especially the constraints on search result diversification

References

- R. A. Bradley and M. E. Terry. The rank analysis of incomplete block designs — I. The method of paired comparisons. *Biometrika*, 39:324–345, 1952.
- J. Carbonell and J. Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. SIGIR ’98, pages 335–336, 1998.
- Z. Dou, S. Hu, K. Chen, R. Song, and J.-R. Wen. Multi-dimensional search result diversification. WSDM ’11, pages 475–484, 2011.
- F. Radlinski and S. Dumais. Improving personalized web search using result diversification. SIGIR ’06, pages 691–692, 2006.

- F. Radlinski and S. Dumais. Improving personalized web search using result diversification. SIGIR ’06, pages 691–692, 2006.

- J. Teevan, S. T. Dumais, and E. Horvitz. Personalizing search via automated analysis of interests and activities. SIGIR ’05, pages 449–456, 2005.
- E. Vee, U. Srivastava, J. Shanmugasundaram, P. Bhat, and S. Yahia. Efficient computation of diverse query results. ICDE ‘08, pages 228–236, 2008.
- F. Wu, J. Xu, H. Li, and X. Jiang. Ranking optimization with constraints. CIKM ‘14, 2014.
- B. Xiang, D. Jiang, J. Pei, X. Sun, E. Chen, H. Li, Context-Aware Ranking in Web Search. SIGIR’10, pages 451-458, 2010.

Thank you!