



CCF YOCSEF Tianjin  
July 11, 2015

# Learning for Search Result Diversification

Jun Xu

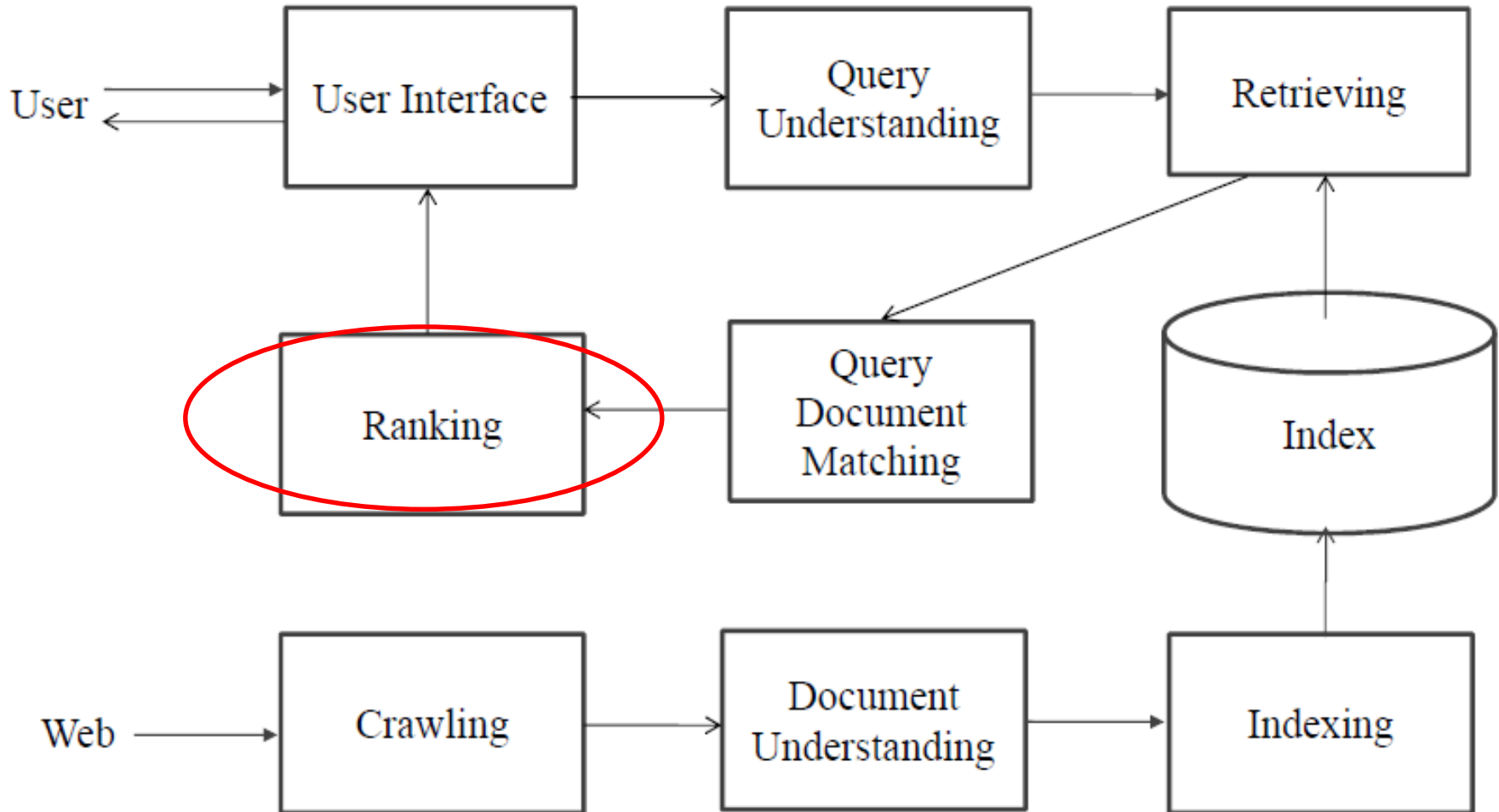
[junxu@ict.ac.cn](mailto:junxu@ict.ac.cn)

Institute of Computing Technology, CAS

# Outline

- Search result diversification
- Learning for search result diversification
- Summary

# Web Search Engine



# Relevance Ranking is Important

nankai university

Web Maps Images News Videos More Search tools

About 1,040,000 results (0.50 seconds)

**南开大学**  
[www.nankai.edu.cn/](http://www.nankai.edu.cn/) Translate this page Nankai University  
包括学校历史, 现状, 教学情况等介绍。  
4.3 ★★★★★ 9 Google reviews · Write a review

94 Weijin Rd, Nankai, Tianjin, China, 300071  
+86 22 2350 4845

**Nankai University - Wikipedia, the free encyclopedia**  
[https://en.wikipedia.org/wiki/Nankai\\_University](https://en.wikipedia.org/wiki/Nankai_University) Wikipedia  
Nankai University (NKU; simplified Chinese: 南开大学; traditional Chinese: 南開大學; pinyin: Nánkāi Dàxué), is a public research university located in Tianjin, ...  
History - Present - Colleges and schools - Rankings

**Study at Nankai University : INTO China**  
[www.intohigher.com/.../universities/nankai...](http://www.intohigher.com/.../universities/nankai...) INTO University Partnerships  
Study in China for a Masters degree taught in English at Nankai University. Nankai is a top ranked university in China for international students and offers ...

**Nankai University | Top Universities**  
[www.topuniversities.com/universities/nan...](http://www.topuniversities.com/universities/nan...) QS World University Rankings  
Is the Nankai University undergraduate program right for you? Read the Top Universities profile.

**Nankai University - China Universities - CUCAS**  
[school.cucas.edu.cn/Study-In-Nankai-University\\_215.html](http://school.cucas.edu.cn/Study-In-Nankai-University_215.html)

- Criteria: relevance
- Ranking model
  - Heuristic: VSM, BM25, LMIR
  - Learning to rank

# Beyond Relevance Ranking

- Freshness
- Response time
- Diversification

Michael Jordan

Web Images News Videos Shopping More Search tools

About 300,000,000 results (0.48 seconds)

**Michael Jordan - Wikipedia, the free encyclopedia**  
[https://en.wikipedia.org/?title=Michael\\_Jordan](https://en.wikipedia.org/?title=Michael_Jordan) - Wikipedia  
Michael Jeffrey Jordan (born February 17, 1963), also known by his initials, MJ, is an American former professional basketball player. He is also an entrepreneur ...  
Jeffrey Jordan - 1984 - Marcus Jordan - Air Jordan

**In the news**

 **NBA 2K16 Special Edition Features Michael Jordan**  
IGN - 11 hours ago  
Basketball legend Michael Jordan will grace the cover for the third time in the video game ...

NBA Draft rumors 2015: Michael Jordan and Phil Jackson may be battling for Frank Kaminsky  
SB Nation - 5 hours ago

LeBron James, Stephen Curry clash most watched NBA Finals since Michael Jordan  
Stuff.co.nz - 1 day ago

[More news for Michael Jordan](#)

**Michael Jordan NBA Stats | Basketball-Reference.com**  
[www.basketball-reference.com](http://www.basketball-reference.com) > Players > J  
Michael Jordan - Career stats, name logs, biographical info, awards, and achievements



**Michael Jordan**  
Basketball player

Michael Jeffrey Jordan, also known by his initials, MJ, is an American former professional basketball player. He is also an entrepreneur, and principal owner and chairman of the Charlotte Hornets. [Wikipedia](#)

**Born:** February 17, 1963 (age 52), Brooklyn, New York City, New York, United States

**Height:** 6' 6" (1.98 m)

**School:** University of North Carolina at Chapel Hill

**Spouse:** Yvette Prieto (m. 2013), Juanita Vanoy (m. 1989–2006)

**Children:** Marcus Jordan, Jeffrey Michael Jordan, Jasmine Mickael Jordan, Ysabel Jordan, Victoria Jordan

# Why Diversification?

Google jaguar

网页 图片 新闻 地图 视频 更多 搜索工具

找到约 443,000,000 条结果 (用时 0.36 秒)

**捷豹中文官方网站 - jaguar.com.cn**  
www.jaguar.com.cn/  
全铝航空科技,轻量化2.0升4涡轮增压发动机或3.0升V6机械增压发动机,更显灵动从容

相关搜索: jaguar汽车价格 jaguar 价格

**捷豹汽车\_最具生命力的英伦汽车-Jaguar中国官方网站**  
www.jaguar.com.cn/  
历经时代变迁,捷豹(jaguar)始终致力于为冰冷的机械赋予灵魂,有无可复制的生命力的极致奢华,捷豹汽车,创新精神,诱人视觉,卓越性能,全球领先

**【捷豹Jaguar】汽车最新资讯 图片 新车上市**  
newcar.xcar.com.cn/  
捷豹汽车: 豪车汽车,汽车图片,捷豹汽车视频,以捷豹XF - 捷豹XJ - 捷豹

**Jaguar**  
www.jaguar.com.cn/  
Official worldwide website for specific markets

**jaguar的**  
八柱一陈政  
中  
中  
【捷  
car.b  
凤凰  
价格  
【进  
car.b  
进口  
进口  
【捷  
car.b  
捷豹  
1922

**捷豹 - 维基百科,自由的百科全书**  
zh.wikipedia.org/zh-cn/捷豹 转为简体网页  
捷豹汽车有限公司(英语:Jaguar Cars Limited)是英国的一家豪华汽车生产商,总部起初座落于英格兰考文垂的布兰兰,后迁至考文垂的惠特利。1922年成立之初製造...

**捷豹汽车|捷豹汽车报价及图片-网上车市**  
product.cheshi.com,汽车大全  
捷豹汽车的历史源远流长,可以追溯到1922年威廉·里昂斯爵士创造出第一辆摩托车跨斗之时。1932年,“捷豹(Jaguar)”的名字首次随着一款完全独自设计制造的...

**捷豹 - 汽车 - 网易**  
product.auto.163.com/brand/1711.html  
捷豹(jaguar)是捷豹汽车集团旗下品牌,品牌起源于英国。捷豹品牌热门车型包括捷豹XF、捷豹XJ、捷豹F-TYPE、捷豹XKR等。网易汽车为您提供捷豹全车型、最新...



Diverse user interests

Google 克里米亚公投

网页 地图 新闻 图片 视频 更多 搜索工具

找到约 3,920,000 条结果 (用时 0.16 秒)

**2014年克里米亚归属公投 - 维基百科,自由的百科全书**  
zh.wikipedia.org/zh-cn/2014年克里米亚归属公投 转为简体网页  
2014年克里米亚归属公投,是克里米亚自治共和国政府於2014年3月16日發起的一場公投,讓克里米亞選民決定是否從烏克蘭獨立,並且加入俄羅斯聯邦。参与者包含...  
背景 - 日期 - 公投選項 - 烏克蘭政府「中止公投決議」

**联合国大会通过决议称克里米亚公投无效** | 联合国 | 克里...  
news.ifeng.com  
2014年3月28日  
16日...行的全

**克里米亚公投**  
news.sina.com.cn  
导读: 克里...  
日,普京下令批... 18

**克里米亚公投**  
news.sina.com.cn  
乌克兰东部两州今日公投称  
甘斯克人民共和国“获得通... 详细

**联合国大会决议称克里米亚公投无效 - 新闻中心\_中国**  
news.sina.com.cn, 国际新闻, 乌克兰... 扬专题  
2014年3月28日 - 决议还说,16日在克里米亚自治共和国和塞瓦斯托波尔市举行的全民公投“无效”,“不能成为改变克里米亚自治共和国和塞瓦斯托波尔市地位的基础”。

**美俄博弈克里米亚公投 新闻频道\_央视网(cctv.com)**  
news.cntv.cn/special/video/Crimea/index.shtml  
克里米亚全民公投于3月17日凌晨结束,民调显示:九成以上选民支持加入俄罗斯联邦。奥巴马强调公投违反乌克兰宪法,永远都不会得到美国和国际社会承认。

**乌克兰局势最新消息: 克里米亚独立公投提前至3月30日举行 ...**  
www.guancha.cn/europe/2014\_03\_02\_209886.shtml  
2014年3月2日 - 乌克兰局势最新消息: 综合外媒报道,乌克兰克里米亚自治共和国议会2月27日决定,今年5月就克里米亚的地位举行公民投票。但克里米亚总理阿克...

**联合国大会通过决议: 克里米亚公投无效\_新闻中心\_中国**  
news.china.com.cn, 新闻中心, 国际, 国际滚动  
2014年3月28日 - 乌克兰克里米亚自治共和国16日举行全民公投,近97%的投票者赞成克里米亚加入俄罗斯。18日,俄总统普京在克里姆林宫向克里米亚及塞瓦斯托波...

**克里米亚公投,解读中国的弃权票(陈破空)**  
www.rfi.org/mandarin/pinglun/chenpokong/ke-03182014114959.html  
2014年3月18日 - 3月16日,克里米亚举行公投,结果:近97%的投票者支持克里米亚脱离乌克兰、加入俄罗斯。俄罗斯总统普京(Vladimir Putin)迅速与克里米亚地区...

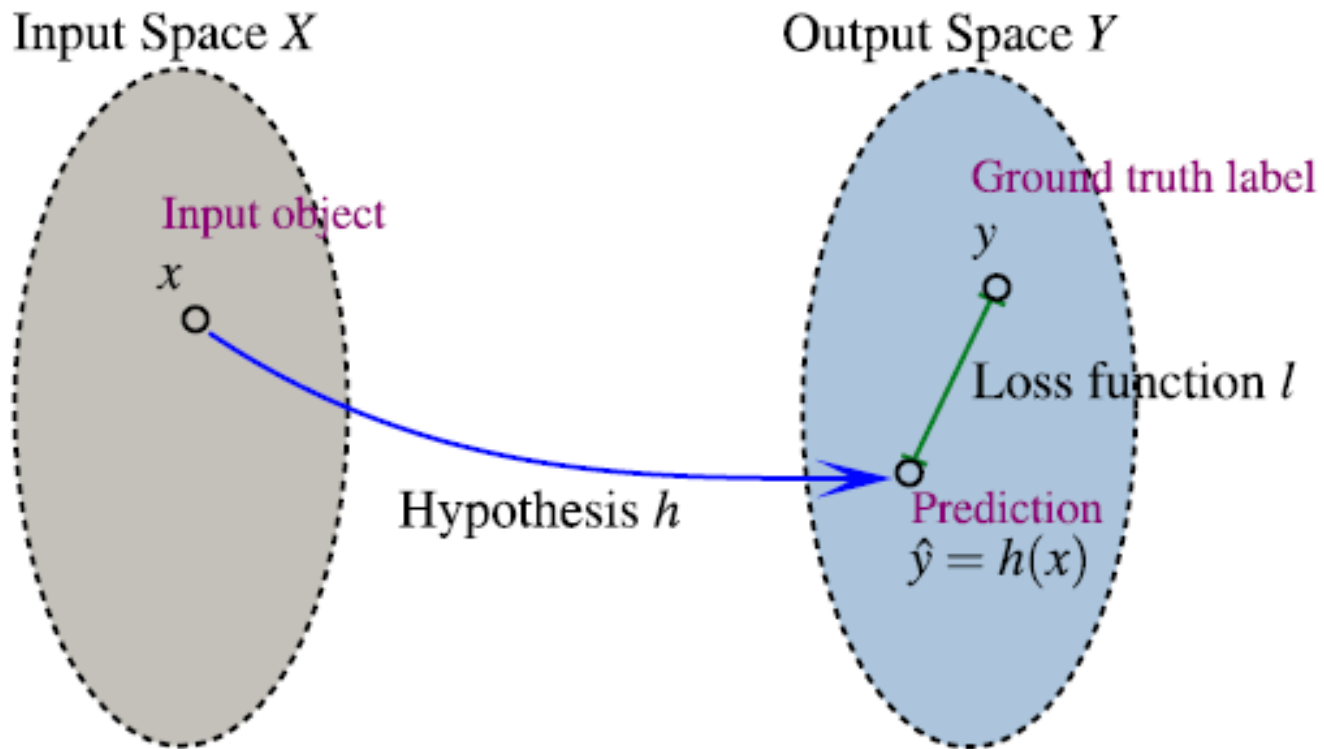
**俄总理首访公投后克里米亚俄决定划为经济特区 - 国际 - 环球网**  
world.huanqiu.com, 国际新闻, 独家  
2014年4月1日 - 梅德韦杰夫访问克里米亚,他是在公投并入俄罗斯后第一个访问该地区的俄罗斯领导人。

Reducing redundant

# Outline

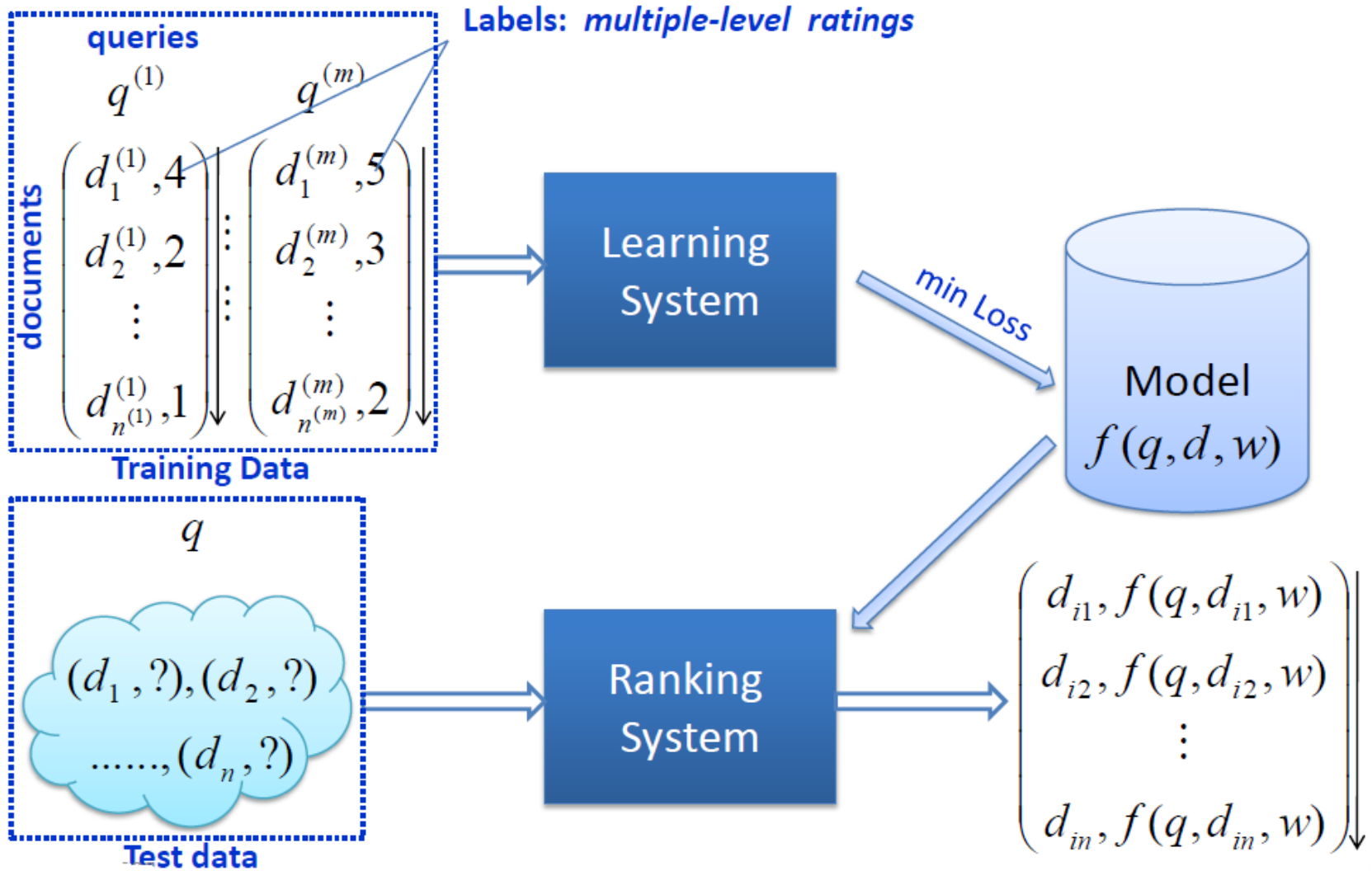
- Search result diversification
- Learning for search result diversification
  - Ranking model
  - Loss function and optimization
  - Experimental results
- Summary

# Framework of Machine Learning





# Bring Machine Learning to Ranking



# Ranking Model for Relevance Learning to Rank

- Independent scoring function, e.g.,  $f(q, d) = \langle \vec{w}, \phi(q, d) \rangle$
- Features  $\phi(q, d)$

Table 1: Relevance Features for learning on ClueWeb09-B collection [21, 19].

Category	Feature Description	Total
$Q-D$	TF-IDF	5
$Q-D$	BM25	5
$Q-D$	QL.DIR	5
$Q-D$	MRF	10
$D$	PageRank	1
$D$	#Inlinks	1
$D$	#Outlinks	1

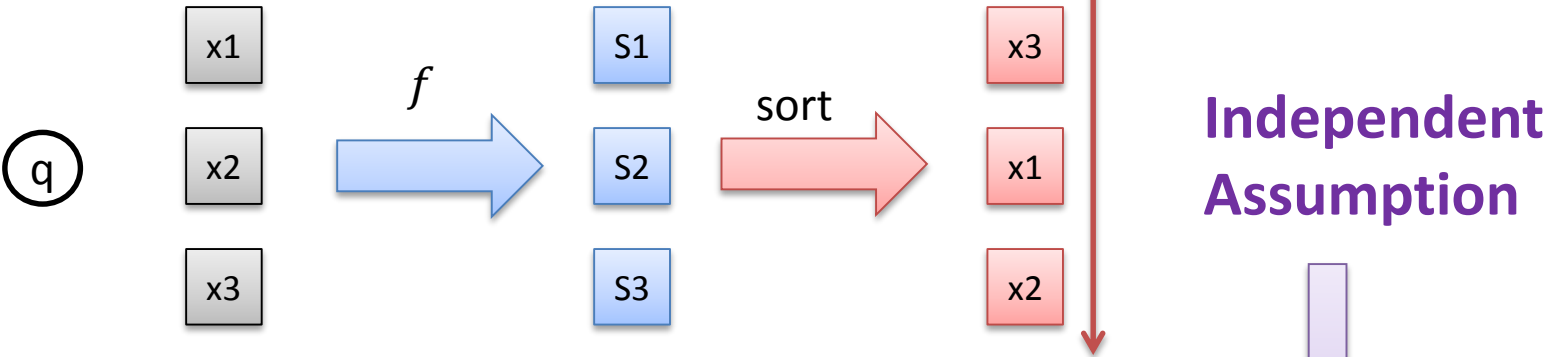
- Ranking: scoring and sorting ascendingly

Query	Documents	score
q	d1	$f(q, d1)$
	d2	$f(q, d2)$
	d3	$f(q, d3)$

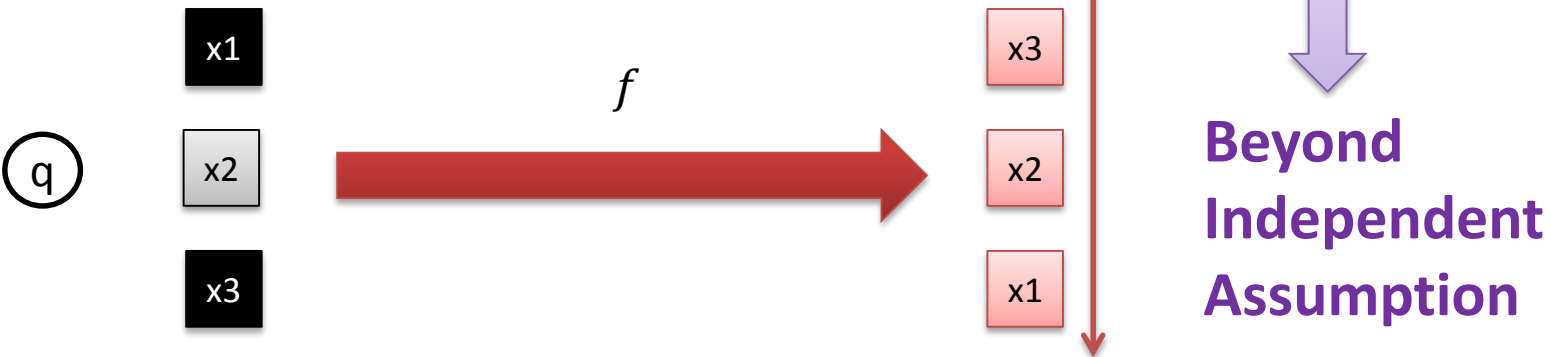


# Beyond Relevance Ranking

- Relevance Ranking



- Diverse Ranking



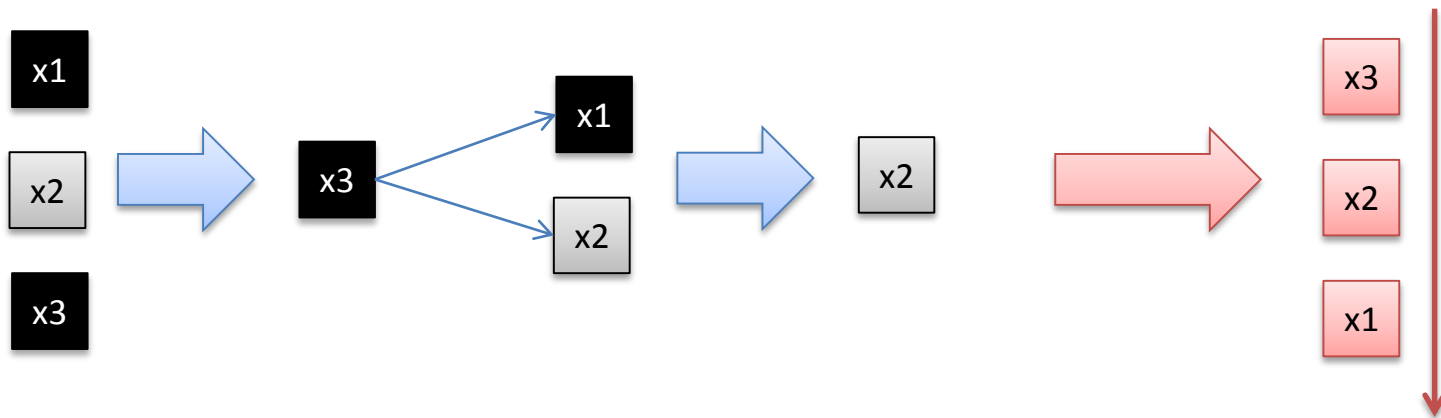
# Maximal Marginal Relevance

- A greedy approach to search result diversification
  - Scoring function

$$\text{MMR} \stackrel{\text{def}}{=} \text{Arg max}_{D_i \in R \setminus S} [\lambda \text{Sim}_1(D_i, Q) - (1 - \lambda) \max_{D_j \in S} \text{Sim}_2(D_i, D_j)]$$

Relevance Relation

- Ranking: sequential selection procedure



# Relational Learning to Rank (R-LTR) Model

- Scoring function for selecting a document

$$f_S(x_i, R_i) = \omega_r^T \mathbf{x}_i + \omega_d^T h_S(R_i), \forall x_i \in X \setminus S$$

Selected documents      Content-based score      Relation-based score

- Ranking with sequential document selection

$$\mathbf{f}(X, R) = (f_{S_0}, f_{S_1}, \dots, f_{S_{n-1}})$$

# Relevance Features $\mathbf{x}_i$

- $f_S(x_i, R_i) = \langle w_r, \mathbf{x}_i \rangle + \langle w_d, h_S(R_i) \rangle, \forall x_i \in X \setminus S$
- Adopting features used in relevance learning to rank
  - Query-Document features
  - Document features

# Relational (Diversity) Features

- $f_S(x_i, R_i) = \langle w_r, x_i \rangle + \langle w_d, h_S(R_i) \rangle, \forall x_i \in X \setminus S$ 
  - $R$ : the relation cube,  $R_{ijk}$  is the  $k^{\text{th}}$  feature describing relation of document  $i$  and document  $j$
  - $R_i$ : the matrix describing the relationship of document  $i$  and other documents

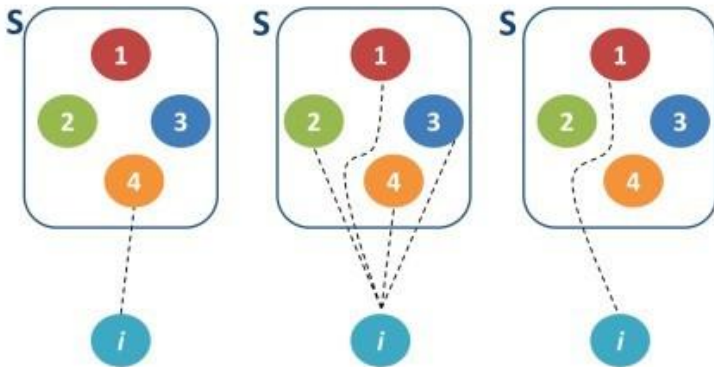
	feature 1	.....	Feature k	...	
d1					
.....					
dj			$R_{ijk}$		
....					

# Definition of $h_S(R_i)$

vector

matrix

- $f_S(x_i, R_i) = \langle w_r, x_i \rangle + \langle w_d, h_S(R_i) \rangle, \forall x_i \in X \setminus S$ 
  - Minimal:  $h_S(R_i) = \left\{ \min_{x_j \in S} R_{ij1}, \dots, \min_{x_j \in S} R_{ijK} \right\}$
  - Mean:  $h_S(R_i) = \left\{ \frac{1}{|S|} \sum_{x_j \in S} R_{ij1}, \dots, \frac{1}{|S|} \sum_{x_j \in S} R_{ijK} \right\}$
  - Maximal:  $h_S(R_i) = \left\{ \max_{x_j \in S} R_{ij1}, \dots, \max_{x_j \in S} R_{ijK} \right\}$



(a) Minimal

(b) Average

(c) Maximal



# Diversity Features

- Feature vector  $[R_{ij1}, R_{ij2}, \dots, R_{ijK}]$ 
  - Subtopic Diversity: based on PLSA  $R_{ij1} = \sqrt{\sum_{k=1}^m (p(z_k|x_i) - p(z_k|x_j))^2}$
  - Text diversity:  $R_{ij2} = 1 - \frac{\mathbf{d}_i \cdot \mathbf{d}_j}{\|\mathbf{d}_i\| \|\mathbf{d}_j\|}$
  - Title diversity
  - Anchor diversity
  - Link-based diversity
  - URL-based diversity
  - ODP-based diversity

# Relevance Ranking Model vs. Relational Ranking Model

## Relevance ranking model

- A single scoring function
- Assigning document scores independently
- Sorting document ascendingly
- Relevance features

## Relational ranking model

- N-1 scoring functions (sharing parameters) for ranking N documents
- Document relations are taken into consideration
- Sequential document selection
- Relevance features + diversity features (d-d)

# Outline

- Search result diversification
- Learning for search result diversification
  - Ranking model
  - Loss function and optimization
  - Experimental results
- Summary

# Relevance Learning to Rank

$$x \rightarrow y$$

## *Pointwise Methods*

- Regression, Order Regression
- OC SVM, McRank

$$(x_1, x_2) \rightarrow y$$

## *Pairwise Methods*

- Pairwise classification
- RankSVM, RankBoost, RankNet, GBRank

$$(x_1, x_2, \dots, x_n) \rightarrow \vec{y}$$

## *Listwise Methods*

- Listwise ranking
- ListMLE, ListNet, RankCosine, StructureSVM, SoftRank, AdaRank

# Conventional Learning to Rank (cont')

- Pointwise

$$\sum_q \sum_{d \in \mathbf{d}} l(f(q, d), y) + \Omega(f)$$

- Pairwise

$$\sum_q \sum_{d >_q d'} l(f(q, d), y, f(q, d'), y') + \Omega(f)$$

- Listwise

$$\sum_{(q, \mathbf{d})} l(\mathbf{F}(q, \mathbf{d}), \mathbf{y}) + \Omega(\mathbf{F})$$

Permutation over  $\mathbf{d}$

# Loss Functions for Search Result Diversification

- Generative approach [Zhu et al., SIGIR '14]
  - Modeling the generation of the result list in a sequential way
  - Using the Plackett-Luce model
  - Optimize with MLE
- Discriminative approach [Xia et al., SIGIR '15]
  - Maximizing margins between “positive” and “negative” rankings
  - Optimize with Perceptron (other methods can also be used)

# Generative Approach [Zhu et al., SIGIR '14]

Process of sequential document selection



Modeling the generation of the result list  
in a sequential way



Loss Function:

Negative log likelihood of generation probability

$$L(\mathbf{f}(X, R), \mathbf{y}) = -\log P(\mathbf{y}|X)$$

$$\begin{aligned} P(\mathbf{y}|X) &= P(x_{y(1)}, x_{y(2)}, \dots, x_{y(n)}|X) \\ &= P(x_{y(1)}|X)P(x_{y(2)}|X \setminus S_1) \cdots P(x_{y(n)}|X \setminus S_{n-1}) \end{aligned}$$

```
1: Initialize  $S_0 \leftarrow \emptyset, \mathbf{y}^{(t)} = (1, \dots, n_t)$ 
2: for  $k = 1, \dots, n_t$  do
3:   bestDoc  $\leftarrow \operatorname{argmax}_{x \in X_t} f_{S_{k-1}}(x, R)$ 
4:    $S_k \leftarrow S_{k-1} \cup \text{bestDoc}$ 
5:    $y^{(t)}(k) \leftarrow$  the index of bestDoc
6: end for
7: return  $\mathbf{y}^{(t)} = (y^{(t)}(1), \dots, y^{(t)}(n_t))$ 
```

# Loss Function of R-LTR

- Plackett-Luce Model

$$\mathbf{P}(\pi | \mathbf{v}) = \prod_{i=1}^M \frac{v_{\pi(i)}}{v_{\pi(i)} + v_{\pi(i+1)} + \dots + v_{\pi(M)}}$$

- Detailed definition

$$P(x_{y(1)} | X) = \frac{\exp\{f_{\phi}(x_{y(1)})\}}{\sum_{k=1}^n \exp\{f_{\phi}(x_{y(k)})\}}, \quad P(x_{y(j)} | X \setminus S_{j-1}) = \frac{\exp\{f_{S_{j-1}}(x_{y(j)}, R_{y(j)})\}}{\sum_{k=j}^n \exp\{f_{S_{k-1}}(x_{y(k)}, R_{y(k)})\}}.$$

- maximize the sum of the likelihood function

$$-\sum_{i=1}^N \sum_{j=1}^{n_i} \log \left\{ \frac{\exp\{\omega_r^T \mathbf{x}_{y(j)}^{(i)} + \omega_d^T h_{S_{j-1}}^{(i)}(R_{y(j)}^{(i)})\}}{\sum_{k=j}^{n_i} \exp\{\omega_r^T \mathbf{x}_{y(k)}^{(i)} + \omega_d^T h_{S_{k-1}}^{(i)}(R_{y(k)}^{(i)})\}} \right\}$$



# Optimization

- Stochastic gradient ascent

$$\Delta\omega_r^{(i)} = \sum_{j=1}^{n_i} \left\{ \frac{\sum_{k=j}^{n_i} \mathbf{x}_{y(k)}^{(i)} \exp\{\omega_r^T \mathbf{x}_{y(k)}^{(i)} + \omega_d^T h_{S_{k-1}^{(i)}}(R_{y(k)}^{(i)})\}}{\sum_{k=j}^{n_i} \exp\{\omega_r^T \mathbf{x}_{y(k)}^{(i)} + \omega_d^T h_{S_{k-1}^{(i)}}(R_{y(k)}^{(i)})\}} - \frac{\mathbf{x}_{y(j)}^{(i)} \exp\{\omega_r^T \mathbf{x}_{y(j)}^{(i)} + \omega_d^T h_{S_{j-1}^{(i)}}(R_{y(j)}^{(i)})\}}{\exp\{\omega_r^T \mathbf{x}_{y(j)}^{(i)} + \omega_d^T h_{S_{j-1}^{(i)}}(R_{y(j)}^{(i)})\}} \right\},$$

$$\Delta\omega_d^{(i)} = \sum_{j=1}^{n_i} \left\{ \frac{\sum_{k=j}^{n_i} h_{S_{k-1}^{(i)}}(R_{y(k)}^{(i)}) \exp\{\omega_r^T \mathbf{x}_{y(k)}^{(i)} + \omega_d^T h_{S_{k-1}^{(i)}}(R_{y(k)}^{(i)})\}}{\sum_{k=j}^{n_i} \exp\{\omega_r^T \mathbf{x}_{y(k)}^{(i)} + \omega_d^T h_{S_{k-1}^{(i)}}(R_{y(k)}^{(i)})\}} - \frac{h_{S_{j-1}^{(i)}}(R_{y(j)}^{(i)}) \exp\{\omega_r^T \mathbf{x}_{y(j)}^{(i)} + \omega_d^T h_{S_{j-1}^{(i)}}(R_{y(j)}^{(i)})\}}{\exp\{\omega_r^T \mathbf{x}_{y(j)}^{(i)} + \omega_d^T h_{S_{j-1}^{(i)}}(R_{y(j)}^{(i)})\}} \right\}.$$

# R-LTR Algorithm

---

## Algorithm 2 Optimization Algorithm

---

**Input:** training data  $\{(X^{(i)}, R^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^N$ ,  
parameter: learning rate  $\eta$ , tolerance rate  $\epsilon$

**Output:** model vector:  $\omega_r, \omega_d$

- 1: Initialize parameter value  $\omega_r, \omega_d$
  - 2: repeat
  - 3:   Shuffle the training data
  - 4:   for  $i = 1, \dots, N$  do
  - 5:     Compute gradient  $\Delta\omega_r^{(i)}$  and  $\Delta\omega_d^{(i)}$
  - 6:     Update model:  $\omega_r = \omega_r - \eta \times \Delta\omega_r^{(i)}$ ,  
                                   $\omega_d = \omega_d - \eta \times \Delta\omega_d^{(i)}$
  - 7:   end for
  - 8:   Calculate likelihood loss on the training set
  - 9: until the change of likelihood loss is below  $\epsilon$
-

# Can R-LTR be Further Improved?

- R-LTR only utilizes “positive” rankings
  - Discriminative learning is effective in many machine learning tasks
- Not all “negative rankings” are equally negative
  - Can be measured with evaluation measures
- There exists a number of diversity evaluation measures such as  $\alpha$ -NDCG, ERR-IA etc.

Directly optimizing performance measures

# Discriminative Approach [Xia et al., SIGIR '15]

$$\min_{\omega_r, \omega_d} \sum_{n=1}^N L(\hat{y}^{(n)}, J^{(n)})$$

$\hat{y}^{(n)}$ : predicted ranking  
 $J^{(n)}$ : ground truth



$$\sum_{n=1}^N \left(1 - E(X^{(n)}, \hat{y}^{(n)}, J^{(n)})\right) \quad E: \text{evaluation measure}$$



Upper bounded

$$\sum_{n=1}^N \max_{\substack{\mathbf{y}^+ \in \mathcal{Y}^+(n); \\ \mathbf{y}^- \in \mathcal{Y}^-(n)}} \left( E(X^{(n)}, \mathbf{y}^+, J^{(n)}) - E(X^{(n)}, \mathbf{y}^-, J^{(n)}) \right)$$

$\mathcal{Y}^+(n)$ : positive rankings

$\mathcal{Y}^-(n)$ : negative rankings

$F$ : ranking model

$$\left[ F(\mathbf{y}^+, X^{(n)}, R^{(n)}) \leq F(\mathbf{y}^-, X^{(n)}, R^{(n)}) \right]$$

$$\hat{y}^{(n)} = \arg \max_{\mathbf{y} \in \mathcal{Y}^{(n)}} F(X^{(n)}, R^{(n)}, \mathbf{y})$$



Upper bounded if  $E \in [0, 1]$

$$\begin{aligned} F(X, R, \mathbf{y}) &= \Pr(\mathbf{y}|X, R) \\ &= \Pr(\mathbf{x}_{y(1)} \cdots \mathbf{x}_{y(M)} | X, R) \\ &= \prod_{r=1}^{M-1} \Pr(\mathbf{x}_{y(r)} | X, S_{r-1}, R) \\ &= \prod_{r=1}^{M-1} \frac{\exp\{f_{S_{r-1}}(\mathbf{x}_i, R_{y(r)})\}}{\sum_{k=r}^M \exp\{f_{S_{r-1}}(\mathbf{x}_i, R_{y(k)})\}} \end{aligned}$$

$$\sum_{n=1}^N \sum_{\mathbf{y}^+; \mathbf{y}^-} \left[ F(X^{(n)}, R^{(n)}, \mathbf{y}^+) - F(X^{(n)}, R^{(n)}, \mathbf{y}^-) \leq E(X^{(n)}, \mathbf{y}^+, J^{(n)}) - E(X^{(n)}, \mathbf{y}^-, J^{(n)}) \right].$$

# Optimization with Perceptron (PAMM)

---

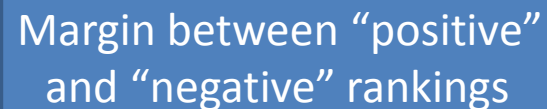
## Algorithm 2 The PAMM Algorithm

---

**Input:** training data  $\{(X^{(n)}, R^{(n)}, J^{(n)})\}_{n=1}^N$ , learning rate  $\eta$ , diversity evaluation measure  $E$ , number of positive rankings per query  $\tau^+$ , number of negative rankings per query  $\tau^-$ .

**Output:** model parameters  $(\omega_r, \omega_d)$

```
1: for  $n = 1$  to  $N$  do
2:    $PR^{(n)} \leftarrow \text{PositiveRankings}(X^{(n)}, J^{(n)}, E, \tau^+)$  {Algorithm 3}
3:    $NR^{(n)} \leftarrow \text{NegativeRankings}(X^{(n)}, J^{(n)}, E, \tau^-)$  {Algorithm 4}
4: end for
5: initialize  $\{\omega_r, \omega_d\} \leftarrow$  random values in  $[0, 1]$ 
6: repeat
7:   for  $n = 1$  to  $N$  do
8:     for all  $\{y^+, y^-\} \in PR^{(n)} \times NR^{(n)}$  do
9:        $\Delta F \leftarrow F(X^{(n)}, R^{(n)}, y^+) - F(X^{(n)}, R^{(n)}, y^-)$ 
          { $F(X, R, y)$  is defined in Equation (8)}
10:      if  $\Delta F \leq E(X^{(n)}, y^+, J^{(n)}) - E(X^{(n)}, y^-, J^{(n)})$ 
          then
11:        calculate  $\nabla\omega_r^{(n)}$  and  $\nabla\omega_d^{(n)}$  {Equation (10)
          and Equation (11)}
12:         $(\omega_r, \omega_d) \leftarrow (\omega_r, \omega_d) + \eta \times (\nabla\omega_r^{(n)}, \nabla\omega_d^{(n)})$ 
13:      end if
14:    end for
15:  end for
16: until convergence
17: return  $(\omega_r, \omega_d)$ 
```



Margin between “positive”  
and “negative” rankings

# Advantages of PAMM

- Online ranking
  - Meets the MMR criteria
- Offline learning
  - Can directly optimizing any diversity evaluation measures
  - Discriminative approach: ability to use both positive rankings and negative rankings
  - Maximizing the margins between positive rankings and negative rankings

# Outline

- Search result diversification
- Learning for search result diversification
  - Ranking model
  - Loss function and optimization
  - Experimental results
- Summary

# Datasets

- TREC datasets: WT2009, WT2010, and WT2011

- Query
  - `<topic number="1" type="faceted">`
  - `<query>obama family tree</query>`
  - `<description>`
    - Find information on President Barack Obama's family history, including genealogy, national origins,
  - `</description>`
  - `<subtopic number="1" type="nav">`
    - Find the TIME magazine photo essay "Barack Obama's Family Tree".
  - `</subtopic>`
  - `<subtopic number="2" type="inf">`
    - Where did Barack Obama's parents and grandparents come from?
  - `</subtopic>`
  - `<subtopic number="3" type="inf">`
    - Find biographical information on Barack Obama's mother.
  - `</subtopic>`
  - `</topic>`

- Label

	Subtopic 1	Subtopic 2	Subtopic 3
Document 1	0	1	0
Document 2	1	1	0
Document 3	0	0	0
.....			



# Evaluation Measures

- $\alpha$ -NDCG

$$\alpha\text{-NDCG}@k = \frac{\sum_{r=1}^k NG(r) / \log(r+1)}{\sum_{r=1}^k NG^*(r) / \log(r+1)},$$

$$NG(r) = \sum_s J(y(r), s) (1 - \alpha)^{C_s(r-1)}$$

$$C_s(r-1) = \sum_{k=1}^{r-1} J(y(k), s)$$

- ERR-IA

$$\text{ERR-IA}@k = \sum_s \Pr(s|q) \text{ERR}@k(s),$$

where  $\text{ERR}@k(s)$  is the expected reciprocal rank score at  $k$  in terms of subtopic  $s$ .

# Experimental Results

**Table 5: Performance comparison of all methods in official TREC diversity measures for WT2009.**

Method	ERR-IA@20	$\alpha$ -NDCG@20
QL	0.164	0.269
ListMLE	0.191(+16.46%)	0.307(+14.13%)
MMR	0.202(+23.17%)	0.308(+14.50%)
xQuAD	0.232(+41.46%)	0.344(+27.88%)
PM-2	0.229(+39.63%)	0.337(+25.28%)
SVM-DIV	0.241(+46.95%)	0.353(+31.23%)
StructSVM( $\alpha$ -NDCG)	0.260(+58.54%)	0.377(+40.15%)
StructSVM(ERR-IA)	0.261(+59.15%)	0.373(+38.66%)
R-LTR	0.271(+65.24%)	0.396(+47.21%)
PAMM( $\alpha$ -NDCG)	0.284(+73.17%)	<b>0.427(+58.74%)</b>
PAMM(ERR-IA)	<b>0.294(+79.26%)</b>	0.422(+56.88%)

**Table 6: Performance comparison of all methods in official TREC diversity measures for WT2010.**

Method	ERR-IA@20	$\alpha$ -NDCG@20
QL	0.198	0.302
ListMLE	0.244(+23.23%)	0.376(+24.50%)
MMR	0.274(+38.38%)	0.404(+33.77%)
xQuAD	0.328(+65.66%)	0.445(+47.35%)
PM-2	0.330(+66.67%)	0.448(+48.34%)
SVM-DIV	0.333(+68.18%)	0.459(+51.99%)
StructSVM( $\alpha$ -NDCG)	0.352(+77.78%)	0.476(+57.62%)
StructSVM(ERR-IA)	0.355(+79.29%)	0.472(+56.29%)
R-LTR	0.365(+84.34%)	0.492(+62.91%)
PAMM( $\alpha$ -NDCG)	0.380(+91.92%)	<b>0.524(+73.51%)</b>
PAMM(ERR-IA)	<b>0.387(+95.45%)</b>	0.511(+69.21%)

**Table 7: Performance comparison of all methods in official TREC diversity measures for WT2011.**

Method	ERR-IA@20	$\alpha$ -NDCG@20
QL	0.352	0.453
ListMLE	0.417(+18.47%)	0.517(+14.13%)
MMR	0.428(+21.59%)	0.530(+17.00%)
xQuAD	0.475(+34.94%)	0.565(+24.72%)
PM-2	0.487(+38.35%)	0.579(+27.81%)
SVM-DIV	0.490(+39.20%)	0.591(+30.46%)
StructSVM( $\alpha$ -NDCG)	0.512(+45.45%)	0.617(+36.20%)
StructSVM(ERR-IA)	0.513(+45.74%)	0.613(+35.32%)
R-LTR	0.539(+53.13%)	0.630(+39.07%)
PAMM( $\alpha$ -NDCG)	0.541(+53.70%)	<b>0.643(+41.94%)</b>
PAMM(ERR-IA)	<b>0.548(+55.68%)</b>	0.637(+40.62%)

- R-LTR outperforms other baselines
- PAMM performs better than R-LTR
- PAMM can directly optimize the evaluation measure

# Effect of MMR

		ranking positions					
		1	2	3	4	5	$\alpha$ -NDCG@5
StructSVM		2, 4	1, 4	2	1, 3	4	0.788
PAMM intermediate rankings	$f_{S_0}$	2, 4	2	4	1, 3	1, 4	0.744
	$f_{S_1}$	2, 4	1, 3	2	4	1, 4	0.803
	$f_{S_2}$	2, 4	1, 3	1, 4	4	2	0.812
	$f_{S_3}$	2, 4	1, 3	1, 4	2	4	0.815

Figure 1: Example rankings from WT2009. Each shaded block represents a document and the number(s) in the block represent the subtopic(s) covered by the document.

# Ability to Improve Evaluation Measures

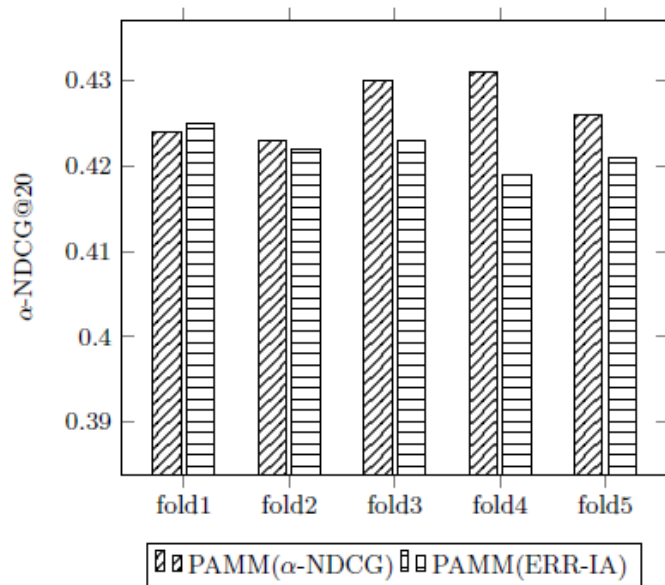


Figure 2: Performance in terms of  $\alpha$ -NDCG@20 when model is trained with  $\alpha$ -NDCG@20 or ERR-IA@20.

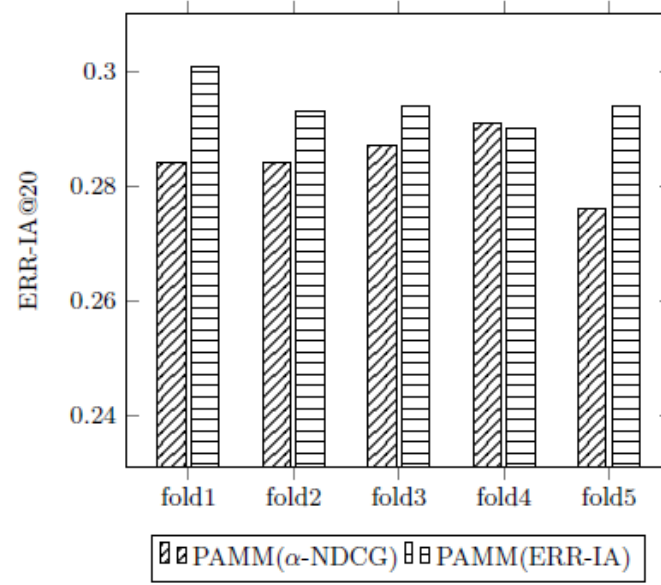


Figure 3: Performance in terms of ERR-IA@20 when model is trained with  $\alpha$ -NDCG@20 or ERR-IA@20.

# Effects of Positive and Negative Rankings

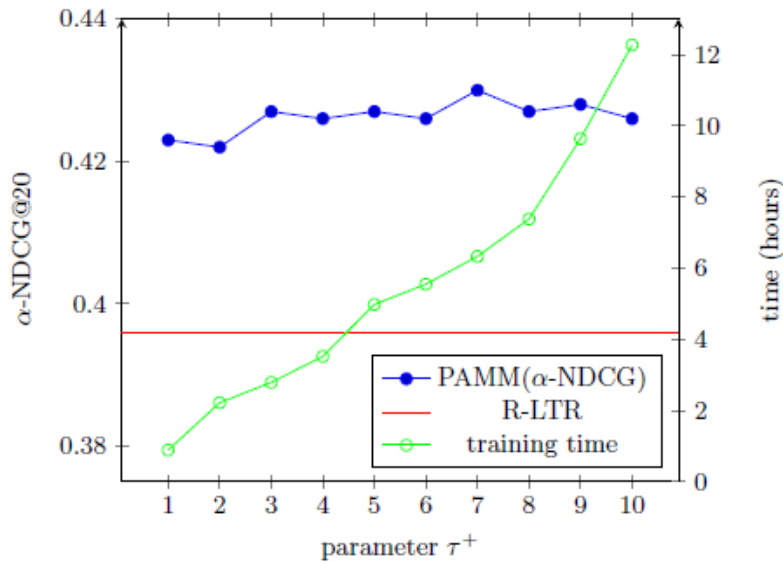


Figure 4: Ranking accuracies and training time w.r.t.  $\tau^+$ .

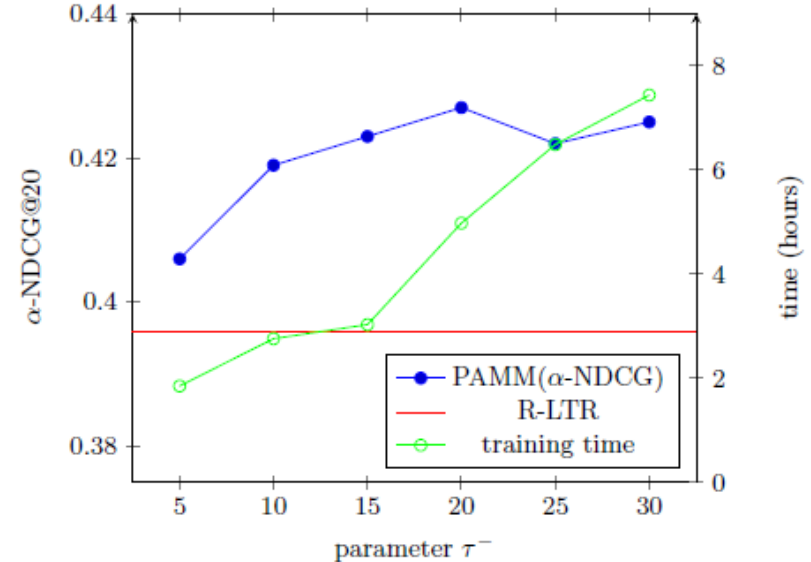


Figure 5: Ranking accuracies and training time w.r.t.  $\tau^-$ .

# Convergence of PAMM

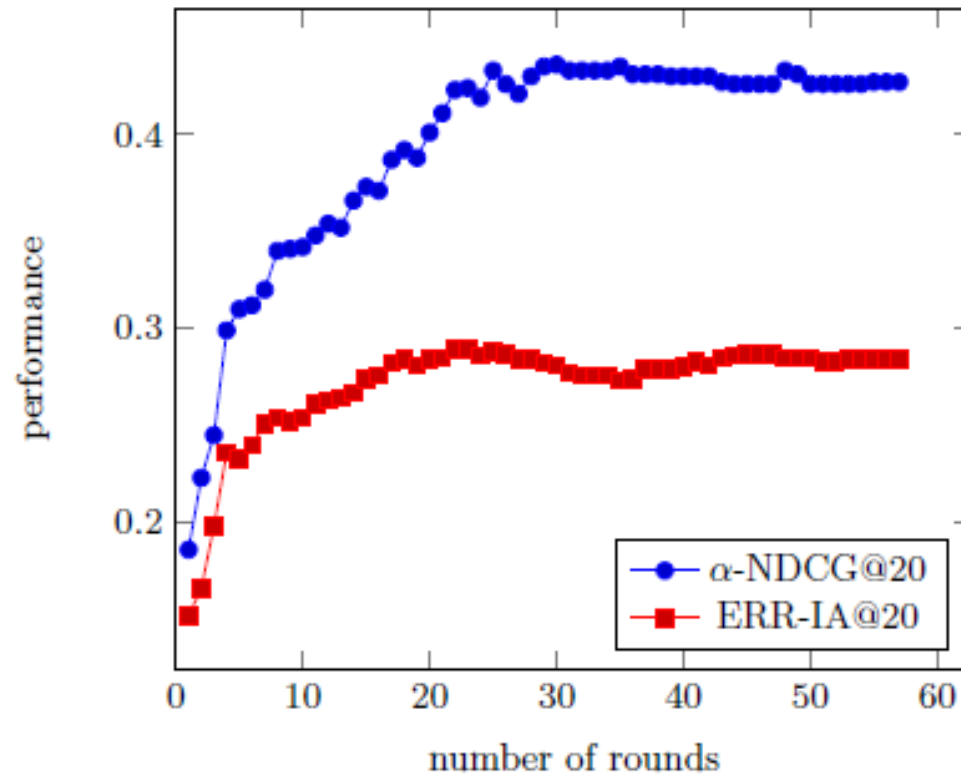


Figure 7: Learning curve of PAMM( $\alpha$ -NDCG).

# Outline

- Search result diversification
- Learning for search result diversification
- **Summary**

# Summary

- New learning to rank models for search result diversification
  - Model: following the MMR criteria
  - Generative learning [Zhu et al., SIGIR '14]
    - Modeling generation process with Luce model
    - Optimizing with MLE
  - Discriminative learning [Xia et al., SIGIR '15]
    - Directly optimizing evaluation measures
    - Utilizing both positive rankings and negative rankings
    - Optimizing with structured Perceptron



# Future Directions

- Learning to rank is not hot in recent years
  - However, a lot of issues not addressed
- New applications: beyond independent relevance
  - Diversification
  - Whole page relevance
  - Topic distillation
- Addressing issues in existing (famous) algorithms
  - E.g., IID assumption in pairwise ranking algorithms such as Ranking SVM does not hold in real world data [Zhang et al., CIKM '15]
- New modeling and optimization tools
  - Deep neural networks?
  - ADMM for large scale learning?

# References

- R. Agrawal, S. Gollapudi, A. Halverson, and S. Jeong. Diversifying search results. WSDM 2009.
- Jaime Carbonell and Jade Goldstein. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. SIGIR 1998.
- O. Chapelle, D. Metzler, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. CIKM 2009.
- C. L. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Buttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. SIGIR 2008.
- T.-Y. Liu. Learning to Rank for Information Retrieval. Springer, 2011.
- **Long Xia, Jun Xu, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. Learning Maximal Marginal Relevance Model via Directly Optimizing Diversity Evaluation Measures. SIGIR 2015.**
- C. X. Zhai, W. W. Cohen, and J. Laerty. Beyond independent relevance: methods and evaluation metrics for subtopic SIGIR, 2003.
- Yaogong Zhang, Jun Xu, Yanyan Lan, Jiafeng Guo, Maoqiang Xie, Yalou Huang, and Xueqi Cheng. Modeling Parameter Interactions in Ranking SVM. CIKM 2015, short paper.
- **Yadong Zhu, Yanyan Lan, Jiafeng Guo, Xueqi Cheng, and Shuzi Niu. Learning for Search Result Diversification. SIGIR 2014.**

# Thanks! Q&A

Homepage:

[www.bigdatalab.ac.cn/~junxu](http://www.bigdatalab.ac.cn/~junxu)