



CCIR 2015  
2015年8月23日  
洛 阳

# 排序学习： 从相关性排序到关系排序

徐 君

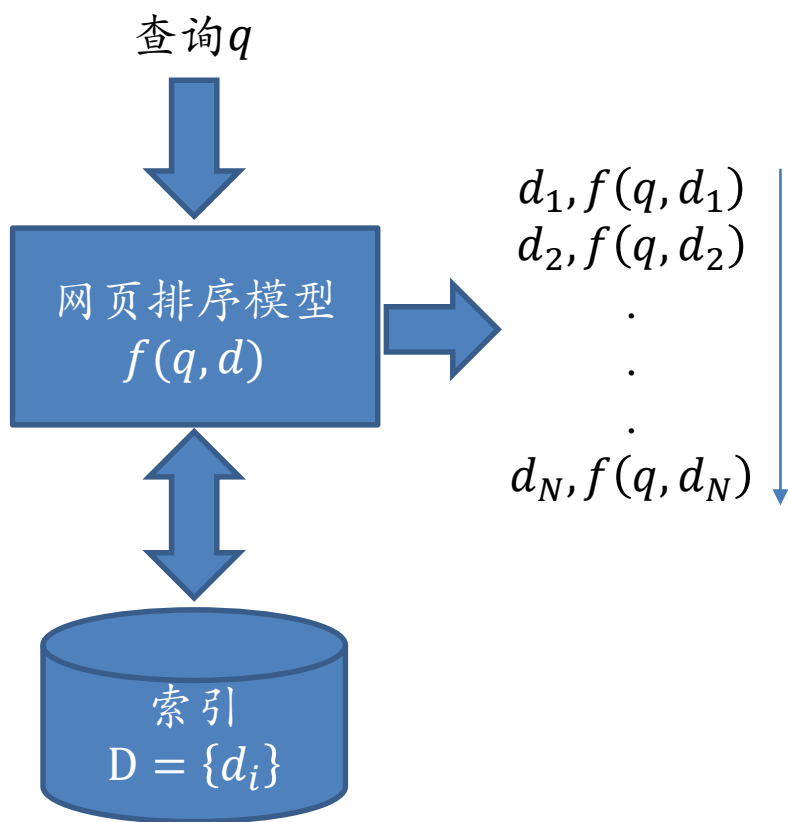
[junxu@ict.ac.cn](mailto:junxu@ict.ac.cn)

中国科学院计算技术研究所

# 提纲

- 相关性排序学习
- 从相关性排序到关系排序
- 总结

# 互联网搜索中的排序



CCIR 2015

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

[第二十一届全国信息检索学术会议\(CCIR2015\)](#)  
CCIR2015青年学者讲坛(8月24日)时间安排及学者介绍 Tutorial A(14:00-15:10):How to generate a good word embedding? 主讲人:刘康 刘康,博士,中科院自动化所...  
[www.ccir2015.com/](#) - 百度快照 - 评价

[第二十一届全国信息检索学术会议\(CCIR2015\)征文通知](#)  
第二十一届全国信息检索学术会议(CCIR2015)开始征稿了,会议官方网站为<http://www.ccir2015.com>,欢迎大家踊跃投稿参会,投稿截止日期为2015年4月30日,更多详...  
[www.cs.sdu.edu.cn/getN...](#) - 百度快照 - 评价

[...和程序委员会主席 - 第二十一届全国信息检索学术会议\(CCIR2015\)](#)  
2015年4月1日 - 2015-04-01 21:23最新消息:CCIR2015指导委员会经过讨论确定:大会主席为清华大学的马少平教授。程序委员会主席为山东大学的马军教授。特向两位主席...  
[ccir2015.com/...jsp?id...](#) - 百度快照 - 评价

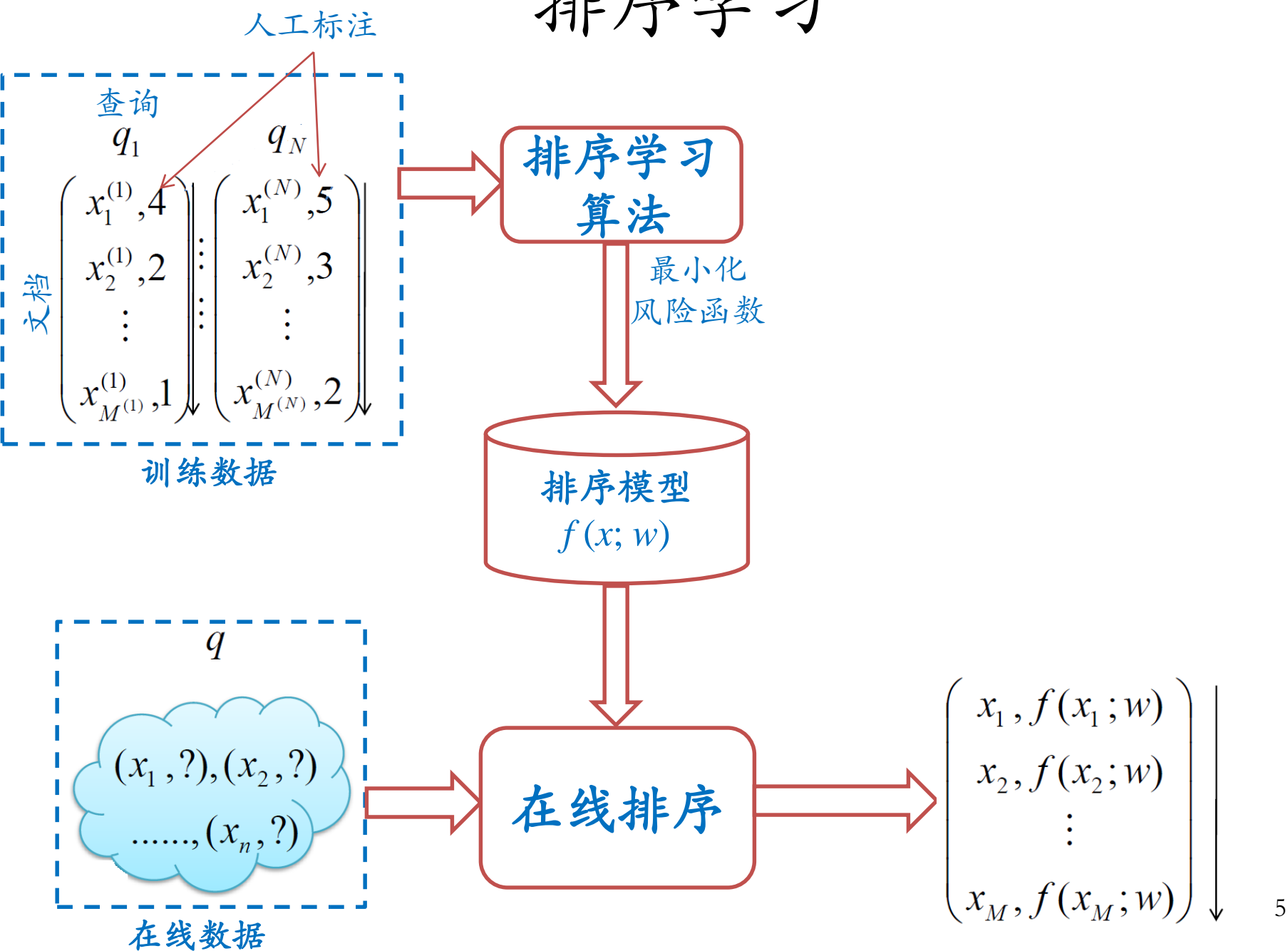
[第二十一届全国信息检索学术会议\(CCIR2015\)](#)  
赞助形式与标准 赞助单位 top ©2015 第二十一届全国信息检索学术会议(CCIR2015) 版权所有  
技术支持:凡科建站电脑版 在线留言 在线地图...  
[m.ccir2015.com/](#) - 百度快照 - 评价

[管理科学与工程学院](#)  
2014年5月8日 - 2 全国信息检索学术会议(CCIR) 由中国计算机学会(CCF)和中国中文信息学会...各刊发表会议“全篇论文”截止日期2015年5月。大会论文投稿咨询: email...  
[einfo.dufe.edu.cn/ind...](#) - 百度快照 - 80%好评

# 排序模型

- 传统信息检索模型
  - 布尔模型(非排序)
  - 向量空间模型、BM25、语言模型(LM4IR)等
  - 词依赖模型(MRF)等
- 基于机器学习的检索模型
  - 排序学习(learning to rank)

# 排序学习



# 相关性排序学习算法

- 模型(打分函数):  $f(q, d)$

- 文档独立打分
- 与位置/上下文无关

- 目标函数

- 基于网页的算法(point-wise)

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

- 基于网页对的算法(pair-wise)

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

- 基于网页列表的算法(list-wise)

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

文档排列

标注列表

# Categorization of the Algorithms

Category	Algorithms
Pointwise Approach	<p><b>Regression:</b> Least Square Retrieval Function (TOIS 1989), Regression Tree for Ordinal Class Prediction (Fundamenta Informaticae, 2000), Subset Ranking using Regression (COLT 2006), ...</p> <p><b>Classification:</b> Discriminative model for IR (SIGIR 2004), McRank (NIPS 2007), ...</p> <p><b>Ordinal regression:</b> Pranking (NIPS 2002), OAP-BPM (EMCL 2003), Ranking with Large Margin Principles (NIPS 2002), Constraint Ordinal Regression (ICML 2005), ...</p>
Pairwise Approach	Learning to Retrieve Information (SCC 1995), Learning to Order Things (NIPS 1998), Ranking SVM (ICANN 1999), RankBoost (JMLR 2003), LDM (SIGIR 2005), RankNet (ICML 2005), Frank (SIGIR 2007), MHR(SIGIR 2007), GBRank (SIGIR 2007), QBRank (NIPS 2007), MPRank (ICML 2007), IRSVM (SIGIR 2006), ...
Listwise Approach	<p><b>Listwise loss minimization:</b> RankCosine (IP&amp;M 2008), ListNet (ICML 2007), ListMLE (ICML 2008), ...</p> <p><b>Direct optimization of IR measure:</b> LambdaRank (NIPS 2006), AdaRank (SIGIR 2007), SVM-MAP (SIGIR 2007), SoftRank (LR4IR 2007), GPRank (LR4IR 2007), CCA (SIGIR 2007), ...</p>

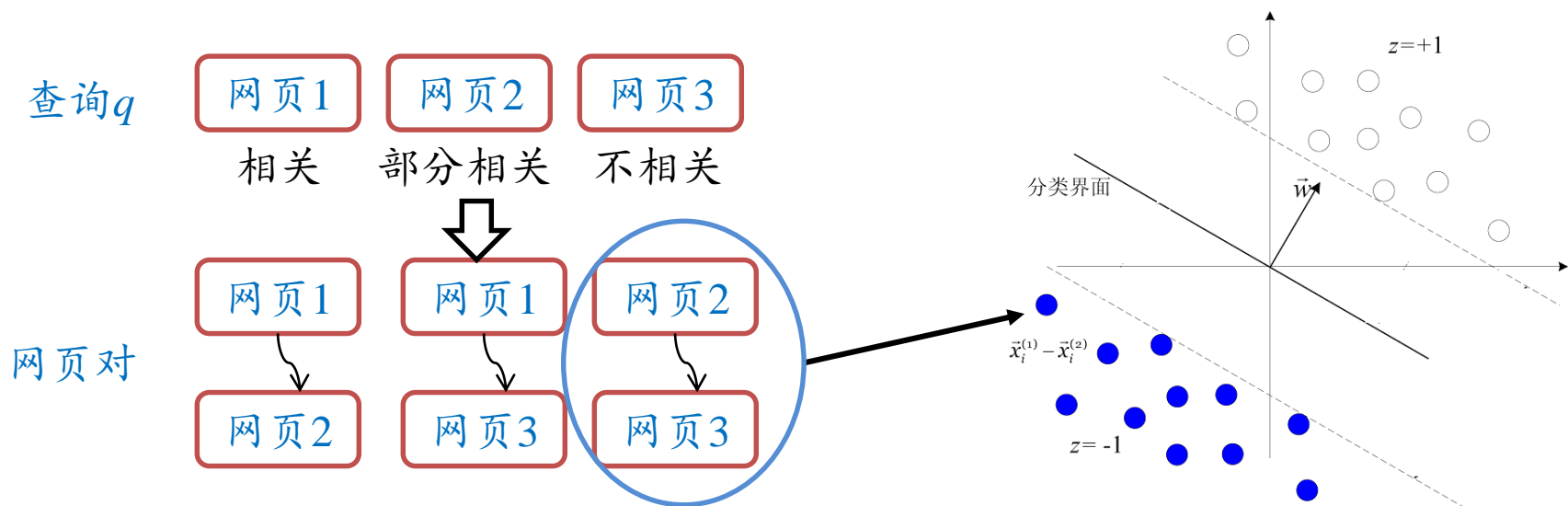
**From Tie-Yan Liu's tutorial slides**

<http://www.contem.org/2010summer/slides/LEARNING%20TO%20RANK%20TUTORIAL%20-%20tyliu.pdf>

# 相关性排序学习进展

—Factorized Ranking SVM [Zhang et al., CIKM '15]

- 网页对排序模型：排序问题  $\rightarrow$  网页对二值分类问题 [Joachims, KDD '02; Freund et al., JMLR '03; Cao et al., SIGIR '06]



## • 潜在问题

- 网页对相互关联，如网页1同时出现在网页对(1,2)和(1,3)中



# Ranking SVM

- Ranking SVM优化问题

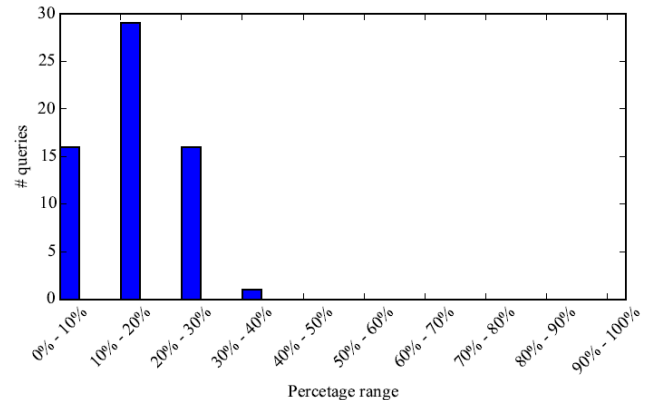
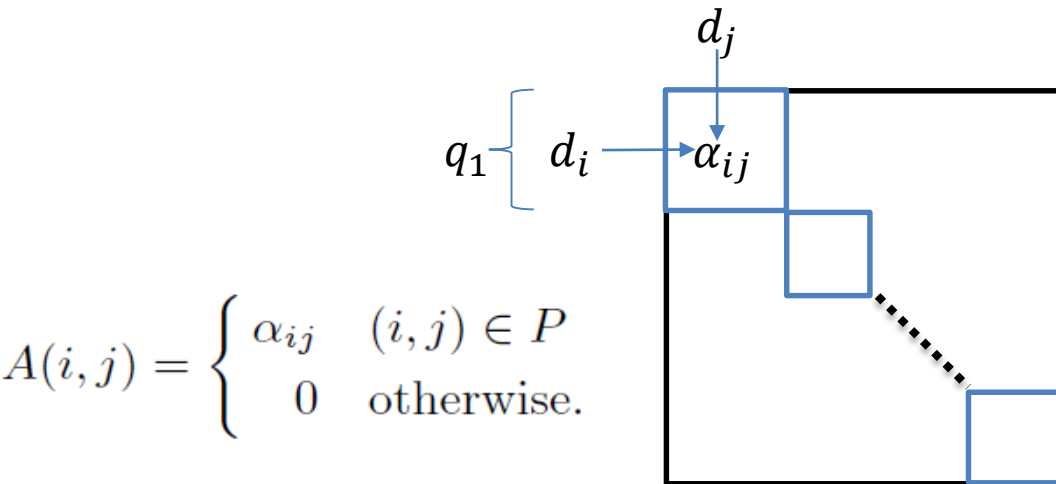
原问题: 
$$\min_{w \in \mathcal{R}^n} \frac{1}{2} \|w\|^2 + C \sum_{(i,j) \in P} [1 - \langle w, x_i - x_j \rangle]_+$$

文档有序对集合

对偶问题: 
$$\min_{\alpha} \frac{1}{2} \alpha^T M \alpha - e^T \alpha,$$
  
 s.t.  $0 \leq \alpha_{ij} \leq C, \forall (i,j) \in P$

$\alpha_{ij}$ 与P中有序对一一对应

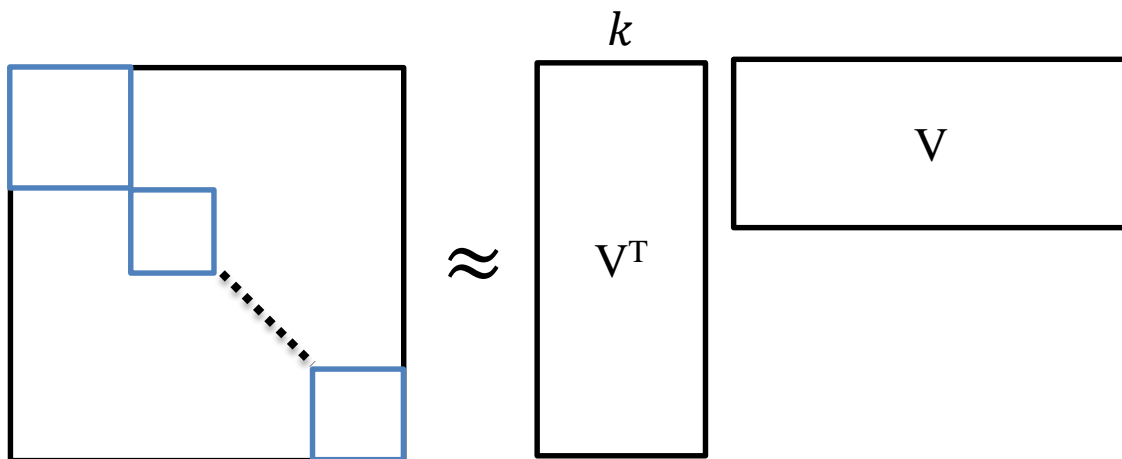
- 对偶变量组成的矩阵A具有**低秩结构**



(a) Capturing 90% energy

# Factorized Ranking SVM

- 改进方案：直接建模低秩结构  $\alpha_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$



- Factorized Ranking SVM 目标函数

$$\min_{\mathbf{v}_1, \dots, \mathbf{v}_N} \frac{1}{2} \left\| \sum_{(i,j) \in P} \langle \mathbf{v}_i, \mathbf{v}_j \rangle (\mathbf{x}_i - \mathbf{x}_j) \right\|^2 + C \sum_{(k,l) \in P} \left[ 1 - \left\langle \sum_{(i,j) \in P} \langle \mathbf{v}_i, \mathbf{v}_j \rangle (\mathbf{x}_i - \mathbf{x}_j), \mathbf{x}_k - \mathbf{x}_l \right\rangle \right]_+$$

优化参数

$\mathbf{w}$

经验风险项

# Factorized Ranking SVM (续)

- 解决了基于网页对排序学习算法中产期存在的理论缺陷，完善了一类算法
- 直接建模网页对之间的交互关系，提升算法的精度
- 极大减少模型的参数规模，从优化 $\alpha_{ij}$ 到优化 $v_{jk}$ ，空间复杂度 $O(N^2) \rightarrow O(KN)$ ，提升了算法的可扩展性

# Factorized Ranking SVM (续)

- 能够改善包括Ranking SVM在内的基于网页对的排序学习算法
- 在网页对比较密集(网页对关联更强)的数据集合上改善更为明显

Table 2: Ranking accuracies on dataset OHSUMED.

Method	MAP	NDCG@1	NDCG@3	NDCG@5
RSVM	0.4427	0.5289	0.4553	0.4392
RankNet	0.404	0.4007	0.3616	0.3388
ListNet	0.4443	0.5134	0.4664	0.4530
Fac-RSVM	<b>0.4463</b>	<b>0.5507</b>	<b>0.4798</b>	<b>0.4546</b>

网页对密集

Table 4: Ranking accuracies on dataset MQ2008.

Method	MAP	NDCG@1	NDCG@3	NDCG@5
RSVM	0.4713	<b>0.3686</b>	0.4277	0.4730
RankNet	0.4522	0.3410	0.3991	0.4500
ListNet	0.4415	0.3244	0.3916	0.4396
Fac-RSVM	<b>0.4714</b>	0.3660	<b>0.4289</b>	<b>0.4731</b>

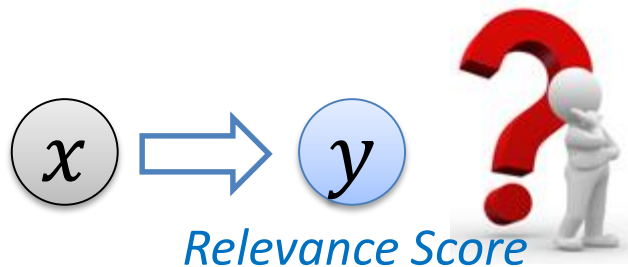
网页对稀疏

# 相关性排序学习进展

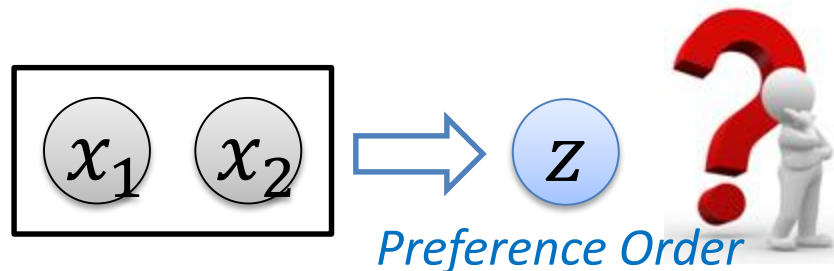
—Top-K排序学习 [Niu et al., SIGIR '12]

- 数据标注过程中的问题

标注网页



标注网页对

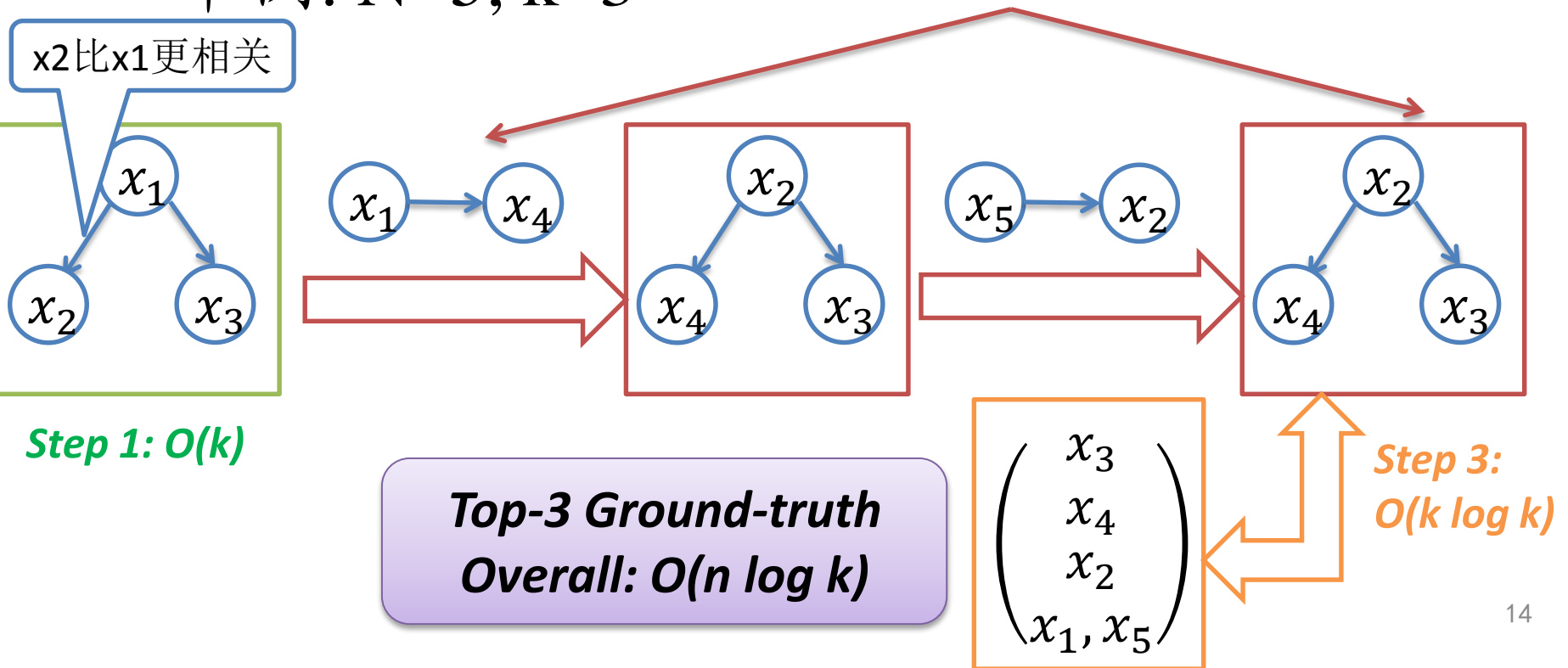


- |                              |                     |
|------------------------------|---------------------|
| (1) 相关性标签如何的确定 (2、3或者5级?)    | (1) 不需要确定标签级别       |
| (2) 标注判断困难 (Perfect or Good) | (2) 标注判断相对简单 (A比B好) |
| (3) 标注结果一致性差                 | (3) 标注结果一致性好        |

**缺点：需要标注的有序对数目庞大**

# Top-K排序学习

- 方法：从N个候选集中标注出最相关的k个文档的全序
  - 有序对标注 + 堆排序
- 举例：N=5, k=3



# 提纲

- 相关性排序学习
- 从相关性排序到关系排序
- 总结

# 相关性排序模型的特点

- 打分函数
  - 文档独立打分
    - 分值与位置无关
    - 分值与其他文档无关
  - 单一打分函数 $f(q, d)$ 
    - 只依赖于查询和待打分文档本身
    - 特征查询-文档匹配特征+文档本身特征
- 排序过程
  - 按打分分值从大到小排序

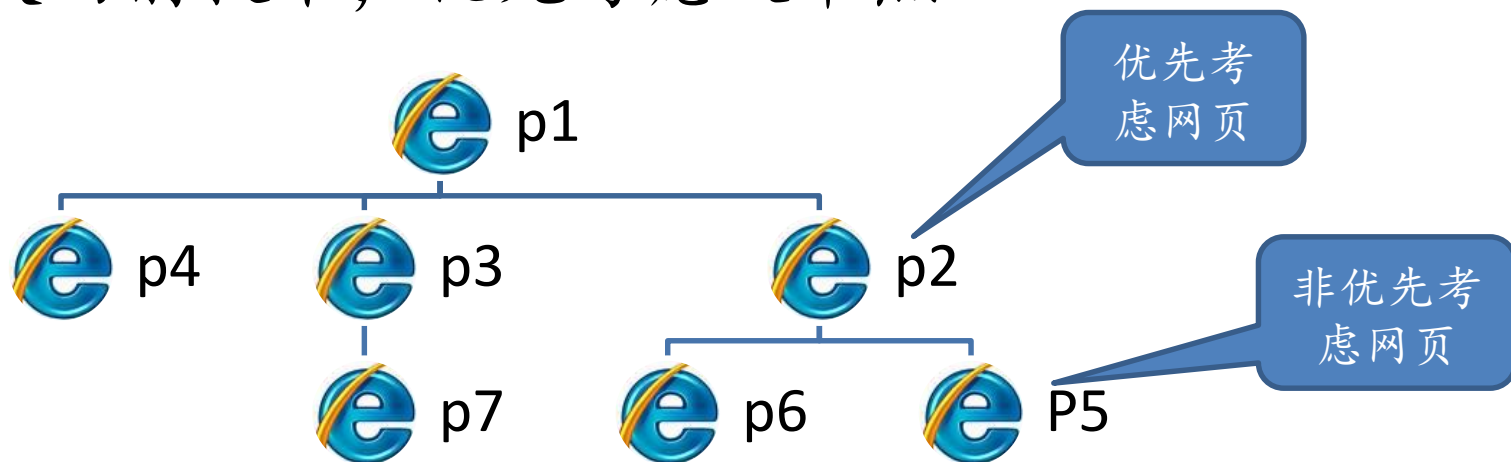


# 相关性排序模型的局限性

- 打分时不考虑网页将要显示的位置，但是
  - 不同网页将显示在不同的排序位置
  - 排序较高的网页可能影响排序较低网页的效应 (搜索结果多样化)
- 网页独立打分
  - 网页间存在多种关系 (相似、链接上下级等)
  - 多个网页将显示在一个结果页面中

# 话题提取 (Topic Distillation)

- Generate a short top-N list, even when a very large set of on-topic documents are available [Craswell & Hawking, TREC '03]
- 在相关的前提下，优先考虑父节点



需要考虑网页间的链接和目录关系

# 搜索结果多样化

## (Search Result Diversification)

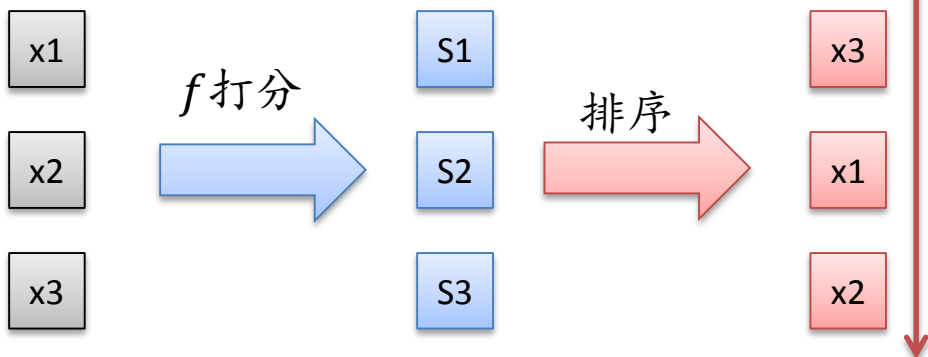
- 搜索结果多样化
  - 用最少的文档覆盖最多的子话题(subtopic)
- 查询 “programming language”

“好”的结果	“坏”的结果
Java	Java
C++	Java
Python	Java

需考虑网页间的相似度

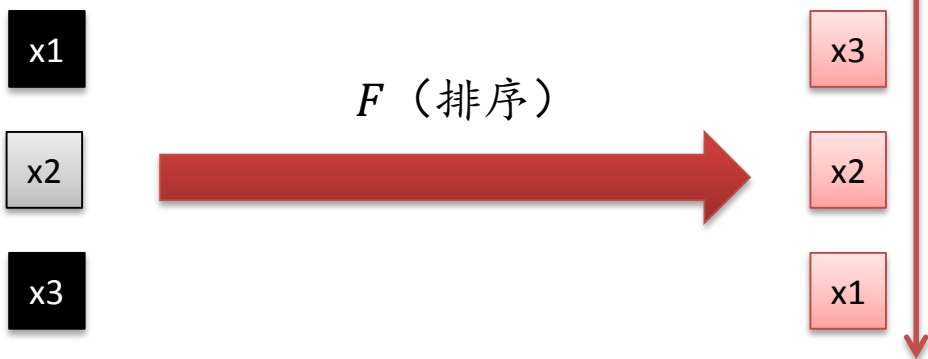
# 相关性排序 vs. 关系排序

- 相关性排序



网页独立性假设

- 关系排序

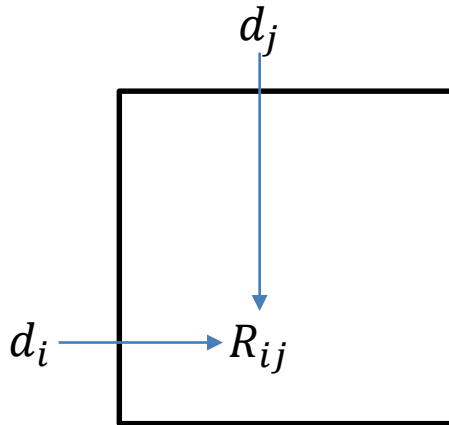


抛弃独立性假设

# 如何表示“关系”

简化为描述所有网页对间的关系

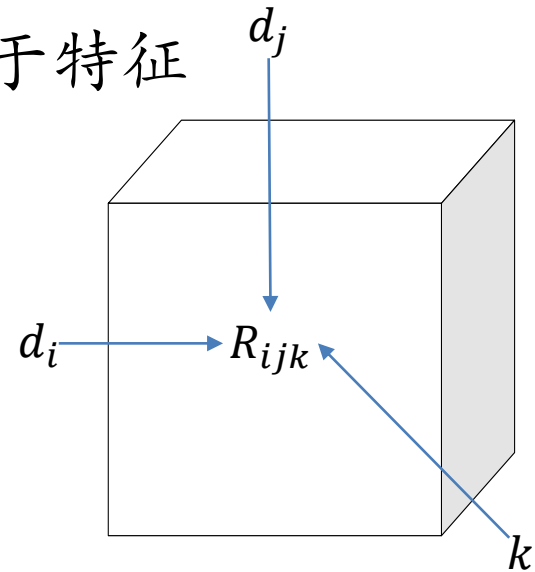
- 静态矩阵



- $R_{ij}$

- $d_i$ 与 $d_j$ 的相似度
- 0-1值, 表示 $d_i$ 与 $d_j$ 的链接关系

- 基于特征  $d_j$



- $R_{ijk}$

- 描述 $d_i$ 与 $d_j$ 关系的第 $k$ 维特征

# 关系排序学习算法

- 通用的关系排序模型
  - Relational RSVM [Qin et al., WWW ‘08]
  - Continuous CRF [Qin et al., NIPS ‘08]
- 搜索结果多样化
  - 基于结构化机器学习模型的改进: [Yue et al. ICML ‘08], [Li et al., WWW ‘09], [Liang et al. KDD ‘14]
  - 关系排序学习: [Zhu et al., SIGIR ‘14] [Xia et al., SIGIR ‘15]

# Relational RSVM—排序模型

[Qin et al., WWW '08]

- 排序过程被定义为一个优化问题

$$f(h(X; \omega), R) = \operatorname{argmin}_z \{l_1(h(X, \omega), z) + \beta l_2(R, z)\}$$

对网页的打  
分序列

相关性效应

网页关系效应

- 应用于话题提取问题

$$l_2(R, z) = \sum_i \sum_j R_{ij} \exp(z_j - z_i) \quad R_{ij} = \begin{cases} 1 & i \text{ is the parent of } j \\ 0 & \text{otherwise} \end{cases}$$



$$f(h(X; \omega), R) = (2I + \beta(2D - R - R^T))^{-1} (2h(X; \omega) - \beta g_1)$$

$$D_{kk} = \frac{1}{2} \left( \sum R_{ik} + \sum R_{ki} \right)$$

# Relational RSVM—排序模型

- 应用于伪相关反馈
  - 相似的文档具有相似的分值

$$l_2(R, z) = \frac{1}{2} \sum_i \sum_j R_{ij} (z_j - z_i)^2 \quad R_{ij} = \text{similarity between } i \text{ and } j$$



$$f(h(X; \omega), R) = (I + \beta(D - R))^{-1} h(X; \omega)$$



# Relational RSVM—优化目标

- 基于Ranking SVM的目标函数

$$\min_{\omega, \xi_q} \frac{1}{2} \omega^T \omega + c \sum_q \mathbf{1}_q^T \xi_q$$

$$\text{s.t. } \forall q \in Q, C_q f(X_q, R_q; \omega) \geq \mathbf{1}_q - \xi_q, \xi_q \geq 0$$

网页对编码

$$C_q = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}$$

话题提取:  $f(X_q, R_q; \omega) = (2I + \beta(2D_q - R_q - R_q^T))^{-1}(2X_q\omega - \beta g_{q,1})$

伪相关反馈:  $f(X_q, R_q; \omega) = (I + \beta(D_q - R_q))^{-1} X_q \omega$

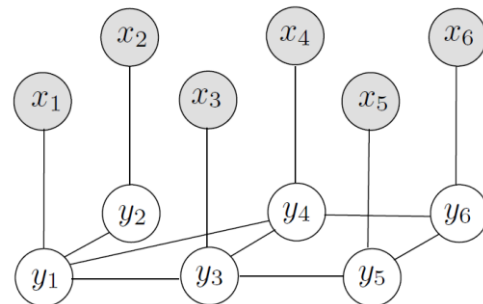
- 存在问题

– 在线: 解优化问题, 实时性差

– 离线: 基于有序对训练, 未能充分利用网页间关系

# Continuous CRF

[Qin et al., NIPS '08]



- 通过CRF描述文档-文档关系

$$\Pr(y^{(q)}|x^{(q)}) = \frac{1}{Z(x^{(q)})} \exp \left\{ \sum_i \sum_{k=1}^{K_1} \alpha_k h_k(y_i^{(q)}, x^{(q)}) + \sum_{i,j} \sum_{k=1}^{K_2} \beta_k g_k(y_i^{(q)}, y_j^{(q)}, x^{(q)}) \right\}$$

相关性

文档间关系

- 应用于伪相关反馈

$$\Pr(y|x) = \frac{1}{Z(x)} \exp \left\{ \sum_i \sum_{k=1}^{K_1} -\alpha_k (y_i - x_{i,k})^2 + \sum_{i,j} -\frac{\beta}{2} S_{i,j} (y_i - y_j)^2 \right\}$$

- 应用于话题提取

$$\Pr(y|x) = \frac{1}{Z(x)} \exp \left\{ \sum_i \sum_{k=1}^{K_1} -\alpha_k (y_i - x_{i,k})^2 + \sum_{i,j} \beta R_{i,j} (y_i - y_j) \right\}$$

# 通用模型存在的问题

- 未能学习文档间的关系
  - 可学习的模型参数 $\omega$ 只刻画了查询-文档相关性
    - 优化目标和可调整参数存在偏差
  - 静态的网页间关系描述
    - 提前计算的相似度
    - 静态链接关系
- 在线生成网页列表时间代价高
  - 需要在线求解优化问题

# 关系排序学习算法

- 通用的关系排序模型
  - Relational RSVM [Qin et al., WWW ‘08]
  - Continuous CRF [Qin et al., NIPS ‘08]
- 搜索结果多样化
  - 基于结构化机器学习模型的改进: [Yue et al. ICML ‘08], [Li et al., WWW ‘09], [Liang et al. KDD ‘14]
  - 关系排序学习: [Zhu et al., SIGIR ‘14] [Xia et al., SIGIR ‘15]

# Structured SVM

[Tsochantaridis et al., ICML '04]

- 预测模型

$$\mathbf{y}^* = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \bar{\mathbf{w}} \cdot \Psi(X, \mathbf{y})$$

列表级别特  
征函数

待排序网页

- 参数学习

$$\frac{1}{2} \|\bar{\mathbf{w}}\|^2 + C \sum_{i=1}^N \xi_i$$

$$\text{s.t. } \bar{\mathbf{w}} \cdot \Psi(X_i, \mathbf{y}_i) - \bar{\mathbf{w}} \cdot \Psi(X_i, \hat{\mathbf{y}}_i) \geq \Delta(\mathbf{y}_i, \hat{\mathbf{y}}_i) - \xi_i$$

for  $1 \leq i \leq N$  and  $\hat{\mathbf{y}}_i \in \mathcal{Y}_i \setminus \{\mathbf{y}_i\}$

对网页的  
排序/打分

$\Delta(\mathbf{y}_i, \hat{\mathbf{y}}_i)$ : 正确答案 $\mathbf{y}_i$ 与不正确答案 $\hat{\mathbf{y}}_i$ 的差异(边界)。可以定义为评价准则间的差异, 从而直接优化评价准则。

- 基于网页列表的排序学习可以看成一种特殊的结构化预测问题

# SVM-DIV

[Yue et al., ICML '08]

- 只关注多样化，不考虑相关性
  - 寻找文档子集，覆盖所有的子话题

$$\min_{\mathbf{w}, \xi \geq 0} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{N} \sum_{i=1}^N \xi_i$$

$$s.t. \forall i, \forall \mathbf{y} \in \mathcal{Y} \setminus \mathbf{y}^{(i)} :$$

$$\mathbf{w}^T \Psi(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) \geq \mathbf{w}^T \Psi(\mathbf{x}^{(i)}, \mathbf{y}) + \Delta(\mathbb{T}^{(i)}, \mathbf{y}) - \xi_i$$

选择的文档

$\mathbb{T}^{(i)} = \{T_1, \dots, T_n\}$ ,  $T_j$  为第  $j$  个文档覆盖的子话题集合

$\mathbf{y}$ : 选择的子文档集合

$\Delta(\mathbb{T}^{(i)}, \mathbf{y})$ :  $\mathbb{T}^{(i)}$  中未被  $\mathbf{y}$  中文档覆盖的子话题的比例

# SVM-DIV—在线文档选择

- $\operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \vec{w}^T \Psi(X, \mathbf{y})$
- 贪心算法

---

Algorithm 1 Greedy subset selection by maximizing weighted word coverage

---

- 1: Input:  $w, x$
- 2: Initialize solution  $\hat{y} \leftarrow \emptyset$
- 3: for  $k = 1, \dots, K$  do
- 4:    $\hat{x} \leftarrow \operatorname{argmax}_{x: x \notin \hat{y}} w^T \Psi(x, \hat{y} \cup \{d\})$
- 5:    $\hat{y} \leftarrow \hat{y} \cup \{\hat{x}\}$
- 6: end for
- 7: return  $\hat{y}$

基于已选择的文档  
逐个添加下一个

# 基于Structured SVM的算法

[Li et al., WWW '09, Liang et al., KDD '14]

- 在Structured SVM优化的基础上，增加将多样化限制条件

$$\min_{\mathbf{w}, \xi \geq 0} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{c}{n} \sum_{i=1}^n \xi_i,$$

subjected to:

$$\forall i, \forall \mathbf{y} \in \mathcal{Y} \setminus \mathbf{y}^{(i)}, \xi_i \geq 0,$$

$$\mathbf{w}^T \Psi(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) \geq \mathbf{w}^T \Psi(\mathbf{x}^{(i)}, \mathbf{y}) + \Delta(\mathbf{y}^{(i)}, \mathbf{y}) - \xi_i.$$

$$\mathbf{w}^T \Psi(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) \geq \sum_{y \in \mathbf{y}^{(i)}} \mathbf{w}^T \Psi(\mathbf{x}^{(i)}, y) - \mu + \xi_i$$

定义为 $\mathbf{y}^{(i)}$ 与 $\mathbf{y}$ 的  
 $\alpha$ -NDCG之差

人工标注的  
正确答案

正确答案中的  
每一个文档

多样化，减少文  
档子话题重复



# 基于Structured SVM的算法(续)

- 贪心策略

---

## Algorithm 2: Greedy subset selection for prediction

---

Input :  $w, \mathbf{x}, u$

1  $\bar{y} \leftarrow \emptyset$

2 for  $k = 1, \dots, K$  do

3      $\bar{x} = \arg \max_{x: x \in \mathbf{x}, x \notin \bar{y}} w^T \Psi(x, u, \bar{y} \cup \{x\})$

4      $\bar{y} \leftarrow \bar{y} \cup \{\bar{x}\}$

5 return  $\bar{y}$

---

# 基于Structured SVM算法存在的问题

- 在线预测与离线训练不一致
  - 训练：假设预测模型为 $\operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \vec{w} \cdot \Psi(X, \mathbf{y})$
  - 预测：贪心策略，局部最优解
- 原因
  - 训练：未利用多样性的特点对 $\mathbf{y}$ 的结构进行合理假设
  - 预测：需要搜索的空间过大，不得已使用贪心策略

# 为多样化排序设计关系排序学习算法

- **模型设计**: 借鉴传统搜索结果多样化模型中的最大化边缘相关度(maximal marginal relevance, MMR)准则和顺序网页选择过程, 对多样化排序的生成过程进行建模
- **参数估计**: 利用最大似然[Zhu et al., SIGIR '14]或者结构感知机[Xia et al., SIGIR '15]进行优化

# 关系排序学习模型设计



# 关系排序学习模型设计

- 打分函数

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

The diagram illustrates the components of the scoring function. A blue box highlights the term  $\omega_r^T \mathbf{x}_i$ , with a blue arrow pointing down to a blue box labeled "相关性打分" (Relevance Scoring). A red box highlights the term  $\omega_d^T h_S(R_i)$ , with a red arrow pointing down to a red box labeled "多样化打分" (Diversity Scoring).

# 关系排序学习模型设计

- 打分函数

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

文档 $x_i$ 与其他文档的关系特征矩阵

关系特征函数

- 排序：顺序文档选择

– 位置 $n$ 对应的打分函数依赖于前 $n-1$ 个位置上所选择的文档 $S$

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

# 相关性特征 $\mathbf{x}_i$

$$f_S(x_i, R_i) = \langle w_r, \mathbf{x}_i \rangle + \langle w_d, h_S(R_i) \rangle, \forall x_i \in X \setminus S$$

- 直接采用相关性排序学习所使用的特征
  - 传统的IR模型/匹配分值: VSM, BM25, LM4IR...
  - 文档长度/PageRank值等
  - .....



# 文档关系特征 $R$

$$f_S(x_i, R_i) = \langle w_r, x_i \rangle + \langle w_d, h_S(R_i) \rangle, \forall x_i \in X \setminus S$$

–  $R: N \times N \times K$ 立方体,  $R_{ijk}$ 为描述文档 $i$ 和文档 $j$ 关系的第 $k$ 维特征

–  $R_i: N \times K$ 矩阵, 描述文档 $i$ 和其他所有文档的关系

	特征1	.....	特征k	.....
文档1				
.....				
文档j			$R_{ijk}$	
...				

# 聚合函数 $h_S(R_i)$

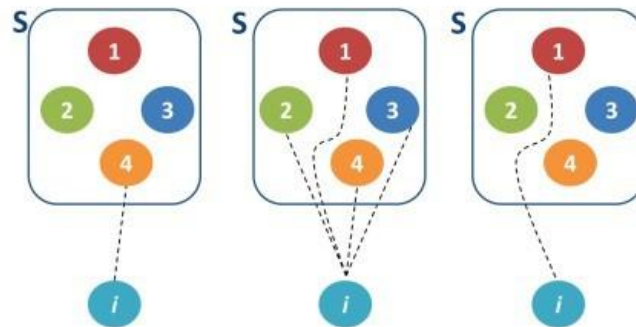
$$f_S(x_i, R_i) = \langle w_r, x_i \rangle + \langle w_d, h_S(R_i) \rangle, \forall x_i \in X \setminus S$$

- 聚合函数，将  $N \times K$  矩阵映射为  $K$  维特征向量

- 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

# 多样化特征

- 通过多个特征刻画文档间的关系(相似度)
- $\phi: D \times D \rightarrow R^K$ 
  - 余弦相似度
  - 映射到话题空间后的余弦相似度
  - 链接相似度
  - URL相似度
  - .....

# 相关性排序学习模型 vs. 关系排序学习模型

## 相关性排序学习

- 单一打分函数
- 文档独立打分，与上下文和位置无关
- 按打分从大到小排序
- 相关性特征

## 关系排序学习

- 对N个待排序文档，N-1个(共享参数的)打分函数
- 打分函数考虑文档间关系和待排序位置
- 顺序文档选择
- 相关性特征+多样化特征(d-d)

# 关系排序学习算法

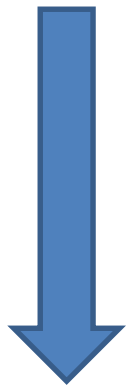
# 目标函数

- 生成式构造目标函数 [Zhu et al., SIGIR '14]
  - 模拟顺序文档选择过程
  - 基于Plackett-Luce 模型对网页排序进行打分
  - 最大似然估计
- 直接优化评价准则 [Xia et al., SIGIR '15]
  - 最大化“正确”排序和“错误”网页排序间的间隔
  - 直接优化评价准则
  - 利用结构感知机算法进行最优化(可采用其他算法)

# 生成式构造目标函数

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

顺序文档选择过程



```
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  $\mathbf{y}^{(t)} = (y^{(t)}(1), \dots, y^{(t)}(n_t))$ 
```

目标函数：生成概率的负对数

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

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

# 生成式构造目标函数

- Luce模型

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

- 基于Luce模型计算每一步选择概率

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

- 目标函数（生成概率负对数）

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



# 目标函数的优化

- 基于随机梯度下降
- 梯度计算公式

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

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



# 算法的优势

- 在机器学习的框架下学习搜索结构多样化模型，同时利用相关性特征和多样化特征
- 利用MMR建模用户自顶向下的浏览行为
- 实验结果

Method	WT2009		WT2010		WT2011	
	ERR-IA@20	$\alpha$ -NDCG@20	ERR-IA@20	$\alpha$ -NDCG@20	ERR-IA@20	$\alpha$ -NDCG@20
QL	0.164	0.269	0.198	0.302	0.352	0.453
ListMLE	0.191	0.307	0.244	0.376	0.417	0.517
MMR	0.202	0.308	0.274	0.404	0.428	0.530
xQuAD	0.232	0.344	0.328	0.445	0.475	0.565
PM-2	0.229	0.337	0.330	0.448	0.487	0.579
SVM-DIV	0.241	0.353	0.333	0.459	0.490	0.591
StructSVM(ERR-IA)	0.261	0.373	0.355	0.472	0.513	0.613
StructSVM( $\alpha$ -NDCG)	0.260	0.377	0.352	0.476	0.512	0.617
R-LTR	0.271	0.396	0.365	0.492	0.539	0.630

# 进一步改进方向

- 只利用了“正确”排序中蕴含的信息
  - 判别模型可以同时利用“正确”和“错误”的排序
- 不同“错误”的排序，其错误的程度不一样
  - 可以利用评价准则对其进行度量
  - 在搜索结果多样化中，常用的评价准则有 $\alpha$ -NDCG, ERR-IA等

在关系排序学习中直接优化评价准则

# 直接优化评价准则

[Xia et al., SIGIR '15]

- 损失函数构造

$\hat{y}^{(n)}$ : 基于多样化排序模型  $f_S$  预测的排序

$$\min_{f_S} \sum_{n=1}^N$$

$$L(\hat{y}^{(n)}, J^{(n)})$$

$J^{(n)}$ : 多样化排序人工标注

$L(\hat{y}^{(n)}, J^{(n)})$ : 当人工标注为  $J^{(n)}$  而预测的排序为  $\hat{y}^{(n)}$  时所产生的“损失”

# 直接优化评价准则

- 在训练数据集上最大化指定的评价准则E

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

由于  $E$  不连续不平滑，难以直接优化。

$E$ : 用户指定的某一排序评价准则，其表示在文档集合 $X$ 上通过模型自动计算出的排序 $y$ 与人工标注的结果 $J$ 之间的一致性程度，常用的评价准则有 $\alpha$ -NDCG和ERR-IA。一般来说 $E \in [0,1]$

# 直接优化评价准则

$$\sum_{n=1}^N \left(1 - E(X^{(n)}, \hat{\mathbf{y}}^{(n)}, J^{(n)})\right)$$



函数上界

$Y^{+(n)}$ : 正确排序( $\alpha$ -NDCG=1)

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

$Y^{-(n)}$ : 错误排序( $\alpha$ -NDCG<1)

$\llbracket \cdot \rrbracket$ : 示性函数

$$\llbracket F(X^{(n)}, R^{(n)}, \mathbf{y}^+) \leq F(X^{(n)}, R^{(n)}, \mathbf{y}^-) \rrbracket$$

$F(X, R, \mathbf{y})$ : 查询级别排序模型

基于Luce模型定义:

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

# 直接优化评价准则

$$\sum_{n=1}^N \left(1 - E(X^{(n)}, \hat{\mathbf{y}}^{(n)}, J^{(n)})\right)$$



$$\sum_{n=1}^N \max_{\substack{\mathbf{y}^+ \in Y^+(n); \\ \mathbf{y}^- \in Y^-(n)}} \left( E(X^{(n)}, \mathbf{y}^+, J^{(n)}) - E(X^{(n)}, \mathbf{y}^-, J^{(n)}) \right) \cdot \llbracket F(X^{(n)}, R^{(n)}, \mathbf{y}^+) \leq F(X^{(n)}, R^{(n)}, \mathbf{y}^-) \rrbracket$$



上界( $E \in [0, 1]$ )

$$\sum_{n=1}^N \sum_{\substack{\mathbf{y}^+ \in Y^+(n); \\ \mathbf{y}^- \in Y^-(n)}} \llbracket F(X^{(n)}, R^{(n)}, \mathbf{y}^+) - F(X^{(n)}, R^{(n)}, \mathbf{y}^-) \leq E(X^{(n)}, \mathbf{y}^+, J^{(n)}) - E(X^{(n)}, \mathbf{y}^-, J^{(n)}) \rrbracket$$



# 基于结构化感知机的优化算法 (PAMM)

- 数据生成：为每一个查询生成正确排序 $\mathbf{y}^+$ 和错误排序 $\mathbf{y}^-$ .
- 参数优化：优化模型参数 $\omega_r$  and  $\omega_d$ .
  - 1:  $\Delta F \leftarrow F(X^{(n)}, R^{(n)}, \mathbf{y}^+) - F(X^{(n)}, R^{(n)}, \mathbf{y}^-)$
  - 2: **if**  $\Delta F \leq E(X^{(n)}, \mathbf{y}^+, J^{(n)}) - E(X^{(n)}, \mathbf{y}^-, J^{(n)})$
  - 3: **then**
  - 4:       calculate  $\nabla \omega_r^{(n)}$  and  $\nabla \omega_d^{(n)}$
  - 5:        $(\omega_r, \omega_d) \leftarrow (\omega_r, \omega_d) + \eta \times (\nabla \omega_r^{(n)}, \nabla \omega_d^{(n)})$
  - 6: **end if**
- 结果返回：**return**  $(\omega_r, \omega_d)$ .

# PAMM优点

- 在机器学习的框架下通过直接优化评价准则，学习最优的搜索结果多样化模型
  - 排序模型建模用户自上而下阅读文档的过程
  - 同时使用了正确排序和错误排序的信息

# 算法效果

WT2009

WT2010

WT2011

Method	ERR-IA@20	$\alpha$ -NDCG@20	ERR-IA@20	$\alpha$ -NDCG@20	ERR-IA@20	$\alpha$ -NDCG@20
QL	0.164	0.269	0.198	0.302	0.352	0.453
ListMLE	0.191	0.307	0.244	0.376	0.417	0.517
MMR	0.202	0.308	0.274	0.404	0.428	0.530
xQuAD	0.232	0.344	0.328	0.445	0.475	0.565
PM-2	0.229	0.337	0.330	0.448	0.487	0.579
SVM-DIV	0.241	0.353	0.333	0.459	0.490	0.591
StructSVM(ERR-IA)	0.261	0.373	0.355	0.472	0.513	0.613
StructSVM( $\alpha$ -NDCG)	0.260	0.377	0.352	0.476	0.512	0.617
R-LTR	0.271	0.396	0.365	0.492	0.539	0.630
PAMM(ERR-IA)	<b>0.294</b>	0.422	<b>0.387</b>	0.511	<b>0.548</b>	0.637
PAMM( $\alpha$ -NDCG)	0.284	<b>0.427</b>	0.380	<b>0.524</b>	0.541	<b>0.643</b>

- R-LTR 优于包括 Structured SVM 在内的其他对照方法
- PAMM 优于 R-LTR
- PAMM 有助于提升选定的评价准则

# 提纲

- 相关性排序学习
- 从相关性排序到关系排序
- 总结

# 总结

- 对排序学习的研究逐渐从相关性排序过渡到关系排序，以解决更加复杂的排序任务
- 相关性排序学习
  - Factorized RSVM
  - Top-K排序
- 关系排序学习
  - 通用的关系排序学习模型
  - 用于搜索结果多样化的关系排序学习模型
    - 基于Structured SVM的改进
    - R-LTR
    - PAMM

# 参考文献

- Yaogong Zhang, Jun Xu, Yanyan Lan, Jiafeng Guo, Maoqiang Xie, Yalou Huang, and Xueqi Cheng. Modeling Parameter Interactions in Ranking SVM. Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM '15).
- T. Joachims, Optimizing Search Engines Using Clickthrough Data, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM, 2002.
- Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. J. Mach. Learn. Res., 4:933-969, 2003.
- Yunbo Cao, Jun Xu, Tie-Yan Liu, Hang Li, Yalou Huang, and Hsiao-Wuen Hon. Adapting ranking SVM to document retrieval. Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '06), Seattle, Washington, USA, pp. 186-193, 2006.
- Shuzi Niu, Jiafeng Guo, Yanyan Lan, Xueqi Cheng, Top-k learning to rank: labeling, ranking and evaluation. Proceedings of the 35th Annual ACM SIGIR conference, Portland, Oregon, USA, August 2012.
- N. Craswell, D. Hawking. Overview of the TREC 2003 Web Track. Proceedings of the TREC 2003. page 78.
- Tao Qin, Tie-Yan Liu, Xu-Dong Zhang, De-Sheng Wang, Wenying Xiong, Hang Li. Learning to Rank Relational Objects and Its Application to Web Search, WWW 2008.
- Tao Qin, Tie-Yan Liu, Xu-Dong Zhang, De-Sheng Wang, Hang Li. Global Ranking Using Continuous Conditional Random Fields, NIPS 2008.

# 参考文献

- I. Tsochantaridis, T. Hofmann, T. Joachims, Y. Altun. Support Vector Machine Learning for Interdependent and Structured Output Spaces. International Conference on Machine Learning (ICML), 2004.
- Yisong Yue, Thorsten Joachims. Predicting Diverse Subsets Using Structural SVMs. International Conference on Machine Learning (ICML), June, 2008.
- Liangda Li, Ke Zhou, Gui-Rong Xue, Hongyuan Zha, Yong Yu. Enhancing diversity, coverage and balance for summarization through structure learning. Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009.
- Shangsong Liang, Zhaochun Ren, Maarten de Rijke. Personalized search result diversification via structured learning. KDD 2014.
- Jaime Carbonell and Jade Goldstein. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. SIGIR 1998.
- Yadong Zhu, Yanyan Lan, Jiafeng Guo, Xueqi Cheng and Shuzi Niu, Learning for Search Result Diversification. Proceedings of the 37th Annual ACM SIGIR Conference, GoldCoast, Australia, 2014. (SIGIR 2014).
- Long Xia, Jun Xu, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. Learning Maximal Marginal Relevance Model via Directly Optimizing Diversity Evaluation Measures. Proceedings of the 38th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '15).

谢谢！

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

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