

# Cross-Domain Learning-to-rank with SVM

Erheng Zhong<sup>1</sup>

<sup>1</sup>Department of Computer Science and Technology, HKUST

COMP621U Presentation, 04/07/2011

# Outline

- 1 Preliminary
  - Ranking
  - Learning-to-rank
  - Transfer Learning
- 2 Cross-domain Learning-to-Rank
  - Motivations
  - Approach: RankSVM
  - Main Results

# Outline

- 1 Preliminary
  - Ranking
    - Learning-to-rank
    - Transfer Learning
- 2 Cross-domain Learning-to-Rank
  - Motivations
  - Approach: RankSVM
  - Main Results

# Definition

- A relationship between a set of items.
- A weak order or total preorder of objects. (mathematics)
- A central part of many information retrieval problems!

# Applications

## ● Search Engine



learning to rank

Search

About 86,200,000 results (0.19 seconds)

Advanced search

Everything

Videos

News

More

All results

Related searches

More search tools

[Learning to rank - Wikipedia, the free encyclopedia](#) - 2 visits - Mar 22

**Learning to rank** or machine-learned ranking (MLR) is a type of supervised or semi-supervised machine learning problem in which the goal is to automatically ...

[Applications - Feature vectors - Evaluation measures - Approaches](#)  
[en.wikipedia.org/wiki/Learning\\_to\\_rank](#) - Cached - Similar

[Yahoo! Learning to Rank Challenge](#) - 3 visits - Mar 23

**Learning to Rank Challenge** is closed! Close competition, innovative ideas, and fierce determination were some of the highlights of the first ever Yahoo! ...

[learningtorankchallenge.yahoo.com/](#) - Cached - Similar

[\[PDF\] Learning to Rank for Information Retrieval This Tutorial](#)

File Format: PDF/Adobe Acrobat - [Quick View](#)

2009/4/12. 1. **Learning to Rank** for Information Retrieval. Tie-Yan Liu. Microsoft Research Asia. A tutorial at WWW 2009. This Tutorial ...  
[www.2009.org/.../TTA-LEARNING%20TO%20RANK%20TUTORIAL.pdf](#) - Similar

[\[PDF\] Learning to Rank](#)

File Format: PDF/Adobe Acrobat - [Quick View](#)

by S. Agarwal - 2005 - [Cited by 5](#) - [Related articles](#)

9 Dec 2005 ... Proceedings of the NIPS 2005 Workshop on. **Learning to Rank**. Edited by. Shivani Agarwal. Massachusetts Institute of Technology ...  
[web.mit.edu/.../Ranking.../proceedings-nips05workshop-ranking.pdf](#) - Similar

[\[PDF\] Learning to Rank using Gradient Descent](#)

File Format: PDF/Adobe Acrobat - [Quick View](#)

**Learning to Rank** using Gradient Descent. Keywords: ranking, gradient descent, neural networks, probabilistic cost functions, internet search. Chris Burges ...  
[research.microsoft.com/pubs/69183/icml\\_ranking.pdf](#)

# Applications

## ● Recommendation System

### What Other Customers Are Looking At Right Now



Getting Away is Deadly: A Mom Zone...  
Sara Rosett  
Kindle Edition  
~~\$0.00~~



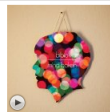
Kindle Wireless Reading Device, Wi-Fi...  
Amazon  
~~\$139.00~~



Invicta Men's 0874 Force Collection...  
~~\$595.00~~ **\$84.99**



Portal 2  
Valve  
PlayStation 3  
~~\$59.99~~ **\$54.99**



Mind Bokeh  
Bibio  
MP3 Download  
~~\$3.99~~

### Digital Cameras Bestsellers



Nikon D3100 14.2MP  
Digital SLR Camera...



Canon PowerShot  
SX130IS 12.1 MP...  
~~\$229.00~~ **\$199.00**



Canon G12 10MP Digital  
Camera with 5x...  
~~\$499.00~~ **\$439.54**



Canon EOS 60D 18 MP  
CMOS Digital SLR...  
~~\$1,299.00~~ **\$1,169.10**



Fujifilm FinePix XP10 12  
MP...  
~~\$149.00~~ **\$109.95**

# Applications

- Computational Advertising

## Ads

### **Fast Gene Synthesis**

Competitive Pricing & 100% Accurate  
Excellent Project Management  
[www.genewiz.com](http://www.genewiz.com)

### **SAS® Data Analysis**

Ensure Your Data is ready for  
Advanced Analytics- SAS Can Help!  
[www.SAS.com](http://www.SAS.com)

### **Biomarker Services**

Gene Expression Biomarker Discovery  
Assay Development + Testing Service  
[www.genebiomarkers.com](http://www.genebiomarkers.com)

### **Yeast Two-Hybrid Service**

Accelerate Your Research!  
Protein Interaction Services

# Outline

- 1 Preliminary
  - Ranking
  - Learning-to-rank
  - Transfer Learning
- 2 Cross-domain Learning-to-Rank
  - Motivations
  - Approach: RankSVM
  - Main Results



# Concepts

Learning-to-rank [1] is to automatically construct a ranking model from training data.

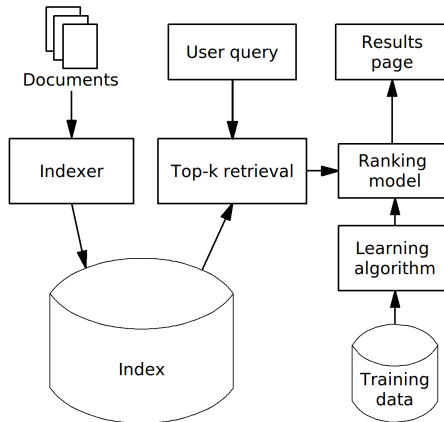
- Training Data:
  - Lists of <query,item> pairs with some partial order specified between pairs
  - $\langle X, \mathbf{y} \rangle$ ; where  $X = \{\mathbf{x}_i = (q_k, t_{kj})\}_{i=1}^{\ell}$  and  $\mathbf{y} = \{y_i\}_{i=1}^{\ell}$
- Ranking Model:
  - A function computing relevance of items for actual queries
  - $f(\mathbf{x} = (q, t)) = \bar{y}$

# Features

<http://research.microsoft.com/en-us/projects/mslr/feature.aspx>

Column in Output	Description
1	TF(Term frequency) of body
2	TF of anchor
3	TF of title
4	TF of URL
5	TF of whole document
6	IDF(Inverse document frequency) of body
7	IDF of anchor
8	IDF of title
9	IDF of URL
10	IDF of whole document
11	TF*IDF of body

# Framework



# Approaches

Three groups with different input representations and loss functions:

- **Pointwise Approach:**

- Each query-document pair in the training data has a numerical or ordinal score.
- A regression problem.

- **Pairwise Approach:**

- A binary classifier which can tell which document is better in a given pair of documents.
- The goal is to minimize average number of inversions in ranking.

- **Listwise Approach:**

- They directly optimize the value of one evaluation measure.

# Approaches

Three groups with different input representations and loss functions:

- **Pointwise Approach:**
  - Each query-document pair in the training data has a numerical or ordinal score.
  - A regression problem.
- **Pairwise Approach:**
  - A binary classifier which can tell which document is better in a given pair of documents.
  - The goal is to minimize average number of inversions in ranking.
- **Listwise Approach:**
  - They directly optimize the value of one evaluation measure.

# Approaches

Three groups with different input representations and loss functions:

- **Pointwise Approach:**
  - Each query-document pair in the training data has a numerical or ordinal score.
  - A regression problem.
- **Pairwise Approach:**
  - A binary classifier which can tell which document is better in a given pair of documents.
  - The goal is to minimize average number of inversions in ranking.
- **Listwise Approach:**
  - They directly optimize the value of one evaluation measure.

# Approaches

Three groups with different input representations and loss functions:

- Pointwise Approach:
  - Each query-document pair in the training data has a numerical or ordinal score.
  - A regression problem.
- Pairwise Approach:
  - A binary classifier which can tell which document is better in a given pair of documents.
  - The goal is to minimize average number of inversions in ranking.
- Listwise Approach:
  - They directly optimize the value of one evaluation measure.

# Outline

- 1 Preliminary
  - Ranking
  - Learning-to-rank
  - **Transfer Learning**
- 2 Cross-domain Learning-to-Rank
  - Motivations
  - Approach: RankSVM
  - Main Results



# Concepts and Notations

Transfer learning [2] refers to the machine learning framework in which one extracts knowledge from some auxiliary domains to help boost the learning performance in a target domain.

- *Auxiliary* domain:  $D_s = \{X_s, \mathbf{y}_s\}$   
*Target* domain:  $D_t = \{X_\ell, \mathbf{y}_\ell; X_u\}$
- $P_s((x), y) \neq P_t((x), y)$

# Approaches

*“what to transfer”* [2]

- Model-based Transfer:

- Discover shared parameters or prior between cross-domain models.

- Feature-based Transfer:

- Find a “good” feature representation that reduces the difference and prediction error between domains.

- Instance-based Transfer:

- Re-weight some labeled data in the auxiliary domain for use in the target domain.

# Approaches

*“what to transfer”* [2]

- Model-based Transfer:
  - Discover shared parameters or prior between cross-domain models.
- Feature-based Transfer:
  - Find a “good” feature representation that reduces the difference and prediction error between domains.
- Instance-based Transfer:
  - Re-weight some labeled data in the auxiliary domain for use in the target domain.

# Approaches

*“what to transfer”* [2]

- Model-based Transfer:
  - Discover shared parameters or prior between cross-domain models.
- Feature-based Transfer:
  - Find a “good” feature representation that reduces the difference and prediction error between domains.
- Instance-based Transfer:
  - Re-weight some labeled data in the auxiliary domain for use in the target domain.

# Approaches

*“what to transfer”* [2]

- Model-based Transfer:
  - Discover shared parameters or prior between cross-domain models.
- Feature-based Transfer:
  - Find a “good” feature representation that reduces the difference and prediction error between domains.
- Instance-based Transfer:
  - Re-weight some labeled data in the auxiliary domain for use in the target domain.

# Applications

- Text classification
- Sentiment analysis
- Image classification
- Name-entity recognition
- WiFi localization
- Spam Filtering
- ...
- **Ranking!**

# Applications

- Text classification
- Sentiment analysis
- Image classification
- Name-entity recognition
- WiFi localization
- Spam Filtering
- ...
- **Ranking!**

# Outline

- 1 Preliminary
  - Ranking
  - Learning-to-rank
  - Transfer Learning
- 2 Cross-domain Learning-to-Rank
  - **Motivations**
  - Approach: RankSVM
  - Main Results



# Sparsity Problem

No enough labeled data in the current domain.

- Heterogeneous feature spaces? Text search  $\Rightarrow$  Image search?
- Out-of-date data? Log data past years  $\Rightarrow$  Search task this year?
- Heterogeneous tasks? Web page ranking  $\Rightarrow$  Expert finding?
- ...

# Outline

- 1 Preliminary
  - Ranking
  - Learning-to-rank
  - Transfer Learning
- 2 Cross-domain Learning-to-Rank
  - Motivations
  - **Approach: RankSVM**
  - Main Results

# Basic RankSVM

RankSVM [3] is a pairwise approach which aims to learn a linear function  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$

$$\begin{aligned} \min_{\mathbf{w}, \xi} \quad & \frac{1}{2} \|\mathbf{w}\|_2^2 + \lambda \sum_{i,j} \xi_{ij} \\ \text{s.t.} \quad & z_{ij} \mathbf{w}^T (\mathbf{x}_i - \mathbf{x}_j) \geq 1 - \xi_{ij}, \quad \xi_{ij} \geq 0, \quad i, j = 1, \dots, \ell \end{aligned} \quad (1)$$

where  $z_{ij}$  is the binary preference defined as follows,

$$z_{ij} = \begin{cases} +1 & \text{if } t_i \succ t_j, \\ -1 & \text{if } t_i \prec t_j. \end{cases}$$

# Model-based Transfer: M-SVM

Schölkopf et al. [4] incorporate knowledge from auxiliary domain using *biased regularization*.

$$\begin{aligned} \min_{w, \xi} \quad & \frac{1}{2} \|w - w_0\|_2^2 + \lambda \sum_{i,j} \xi_{ij} \\ \text{s.t.} \quad & z_{ij} w^T (\mathbf{x}_i^\ell - \mathbf{x}_j^\ell) \geq 1 - \xi_{ij}, \quad \xi_{ij} \geq 0, \quad i, j = 1, \dots, \ell \end{aligned} \tag{2}$$

# Instance-based Transfer: I-SVM

Chen et al. [5] pick those relevant instances from auxiliary domain and eliminate others, by adding weights for instances in the auxiliary domain.

$$\begin{aligned} \min_{w, \xi, \xi^0} \quad & \frac{1}{2} \|w\|_2^2 + \lambda \sum_{i,j} \xi_{ij} + \lambda \sum_{i,j} \rho_{ij} \xi_{ij}^0 \\ \text{s.t.} \quad & z_{ij} w^T (\mathbf{x}_i^\ell - \mathbf{x}_j^\ell) \geq 1 - \xi_{ij}, \quad \xi_{ij} \geq 0, \quad i, j = 1, \dots, \ell \\ & z_{ij}^0 w^T (\mathbf{x}_i^s - \mathbf{x}_j^s) \geq 1 - \xi_{ij}^0, \quad \xi_{ij}^0 \geq 0, \quad i, j = 1, \dots, s \end{aligned} \quad (3)$$

where  $\rho_{ij}$  is the weight on the labeled data pairs in the auxiliary domain.

# Feature-based Transfer: F-SVM

Chen et al. [5] transform instances into a common feature space by learning a projection matrix  $\theta \in \mathbb{R}^{d \times d}$

$$\begin{aligned} \min_{w, \xi, \xi^0, \theta} \quad & \frac{1}{2} \|w\|_2^2 + \lambda \sum_{i,j} \xi_{ij} + \lambda \sum_{i,j} \xi_{ij}^0 \\ \text{s.t.} \quad & z_{ij} w^T \theta^T (\mathbf{x}_i^\ell - \mathbf{x}_j^\ell) \geq 1 - \xi_{ij}, \quad \xi_{ij} \geq 0, \quad i, j = 1, \dots, \ell \\ & z_{ij}^0 w^T \theta^T (\mathbf{x}_i^s - \mathbf{x}_j^s) \geq 1 - \xi_{ij}^0, \quad \xi_{ij}^0 \geq 0, \quad i, j = 1, \dots, s \end{aligned} \tag{4}$$

# Outline

- 1 Preliminary
  - Ranking
  - Learning-to-rank
  - Transfer Learning
- 2 Cross-domain Learning-to-Rank
  - Motivations
  - Approach: RankSVM
  - Main Results

# Measures

- NDCG (Normalized Discounted Cumulative Gain)
- MAP (Mean Average Precision)



# Datasets: Model-based Transfer

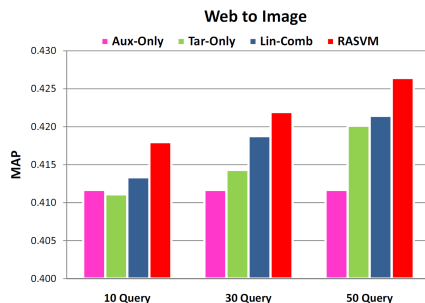
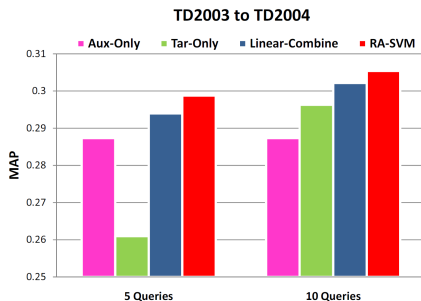
**Table 1: Ranking Adaptation Dataset Information.**

Dataset	#Query	#Query-Document	Relevance Degree	Feature Dimension
TD2003	50	49171	2	44
TD2004	75	74170	2	44
Web Page Search	2625	122815	5	354
Image Search	2053	100404	3	354

**Table 2: Ranking Adaptation Experiment Settings.**

Auxiliary Domain	Train	Validate	Test
TD2003	30	-	20
Web Page Search	500	-	2125
Target Domain	Adapt Pool	Validate	Test
TD2004	30	5	30
Image search	500	10	1543

# Results: Model-based Transfer



# Datasets: Feature-based and Instance-based Transfer

**Table 1** The usage of datasets for cross domain learning to rank

Group	Source domain	Target domain	No. query of $D_s$	No. query of $D_t \cup T$
1	AP	OHSUMED	150	106
2	WSJ	AP	126	150
3	WSJ	OHSUMED	126	106
4	td2003	td2004	50	75
5	hp2003	hp2004	150	75
6	np2003	np2004	150	75

# Results: Instance-based and Feature-based Transfer

**Table 3** Comparison on MAP values (ratio = 0.1, 5–15 queries in target domain)

Group	Source domain	Target domain	LRank <sub>std</sub>	LRank <sub>mix</sub>	LRank <sub>mix_w</sub>	CLRank <sub>feat</sub>	CLRank <sub>ins</sub>
1	AP	OHSUMED	0.284	0.267	0.291	<b>0.320</b> (12.7%)	0.257 (−9.5%)
2	WSJ	AP	0.355	0.361	0.370	<b>0.391</b> (10.1%)	0.359 (1.1%)
3	WSJ	OHSUMED	0.285	0.271	0.293	<b>0.309</b> (8.4%)	0.278 (−2.5%)
4	td2003	td2004	0.178	0.157	0.183	0.185 (3.9%)	<b>0.190</b> (6.7%)
5	hp2003	hp2004	0.647	0.644	0.664	<b>0.690</b> (6.6%)	0.655 (1.2%)
6	np2003	np2004	0.504	0.496	0.536	<b>0.569</b> (12.9%)	0.566 (12.3%)

**Table 4** Comparison on NDCG@5 values (ratio = 0.1, 5–15 queries in target domain)

Group	Source domain	Target domain	LRank <sub>std</sub>	LRank <sub>mix</sub>	LRank <sub>mix_w</sub>	CLRank <sub>feat</sub>	CLRank <sub>ins</sub>
1	AP	OHSUMED	0.399	0.364	0.417	<b>0.445</b> (11.5%)	0.365 (−8.5%)
2	WSJ	AP	0.679	0.689	0.699	<b>0.726</b> (6.9%)	0.684 (0.7%)
3	WSJ	OHSUMED	0.418	0.357	0.425	<b>0.438</b> (4.8%)	0.408 (−2.4%)
4	td2003	td2004	0.258	0.217	0.266	0.268 (3.9%)	<b>0.275</b> (6.6%)
5	hp2003	hp2004	0.679	0.689	0.702	<b>0.726</b> (6.9%)	0.684 (0.7%)
6	np2003	np2004	0.532	0.532	0.559	<b>0.599</b> (12.6%)	0.591 (11.1%)

# Summary

- M-SVM: *Adapt* a trained model to fit the data in the target domain.
- F-SVM: *Transform* the feature space to well bridge auxiliary and target domains.
- I-SVM: *Leverage* relevant instances in the auxiliary domains to increase the training data pool in the target domain.

**M-SVM is efficient while F-SVM and I-SVM are flexible.**

# Summary

Methods	Pointwise	Pairwise	Listwise
Model-based	-	✓	✓
Feature-based	-	✓	✓
Instance-based	-	✓	<b>Y?</b>

# Reference I



Tie-Yan Liu

*Learning to Rank for Information Retrieval.*

Foundations and Trends in Information Retrieval:  
3(3):225-331, 2009



Sinno Jialin Pan and Qiang Yang.

A Survey on Transfer Learning.

*IEEE Transactions on Knowledge and Data Engineering*,  
22(10):1345-1359, 2010.



Thorsten Joachims.

Optimizing Search Engines Using Clickthrough Data.

Proceedings of the ACM SIGKDD International Conference  
on Knowledge Discovery and Data Mining (KDD'2002),  
133-142, 2002.

# Reference II



Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola.  
Ranking Model Adaptation for Domain-specific Search.  
Proceeding of the ACM Conference on Information and  
Knowledge Management (CIKM'2009), 197-206, 2009.



Depin Chen, Yan Xiong, Jun Yan, Gui-Rong Xue, Gang  
Wang, and Zheng Chen.  
Knowledge Transfer for Cross Domain Learning to Rank.  
*Information Retrival*, 13:236-253, 2010.