Cross-Domain Learning-to-rank with SVM

Erheng Zhong[1]

[1] Department of Computer Science and Technology, HKUST

COMP621U Presentation, 04/07/2011

Outline

Outline

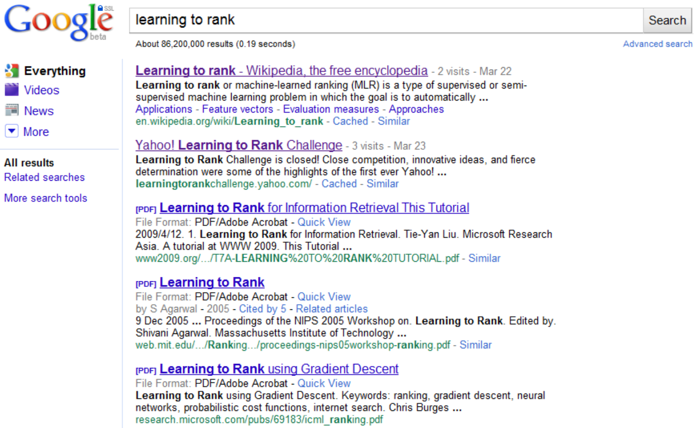
1. Preliminary
   - Ranking
   - Learning-to-rank
   - Transfer Learning
2. Cross-domain Learning-to-Rank
   - Motivations
   - Approach: RankSVM
   - Main Results

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!

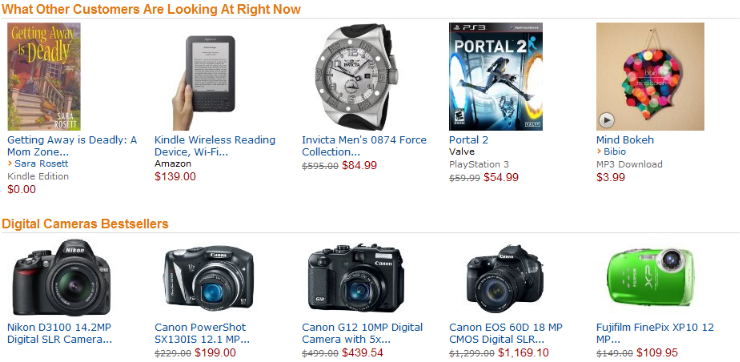
Applications

- Search Engine



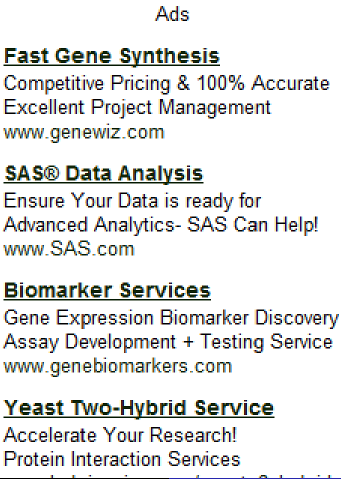
Applications

- Recommendation System



Applications

- Computational Advertising



Outline

Concepts

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

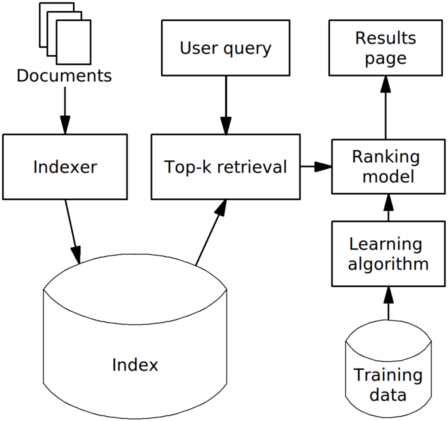
- Training Data:
  - Lists of <query,item> pairs with some partial order specified between pairs
  - $< X, \mathbf{y} >$; where $X = \{\mathbf{x}_i = (q_k, t_{kj})\}_{i=1}^{\ell}$ and $\mathbf{y} = \{y_i\}_{i=1}^{\ell}$
- Ranking Model:
  - A function computing relevance of items for actual queries
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

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


Documents
User query
Results page
Indexer
Top-k retrieval
Ranking model
Learning algorithm
Index
Training data

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

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

Basic RankSVM

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

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

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*.

$$\min_{w,\xi} \quad \frac{1}{2}||w - w_0||_2^2 + \lambda \sum_{i,j} \xi_{ij} \tag{2}$$

$$\text{s.t.} \quad z_{ij}\, w^T(\mathbf{x}_i^\ell - \mathbf{x}_j^\ell) \geq 1 - \xi_{ij},\ \xi_{ij} \geq 0,\ i,j = 1,\ldots,\ell$$

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.

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

where ρ_{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 $\theta \in \mathbb{R}^{d \times d}$

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

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

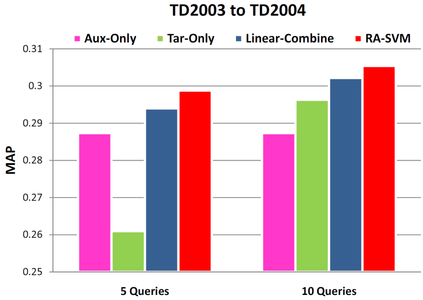
Table 1: Ranking Adaptation Dataset Information.

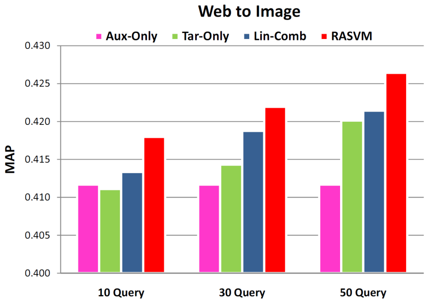
Dataset	#Query	#Query-Document	Relevance Degree	Feature Dimension
TD2003	50	49171	2	44
TD2004	75	74170	2	44
Web Page Search	2625	122815	5	354
Image Search	2053	100404	3	354

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	$LRank_{std}$	$LRank_{mix}$	$LRank_{mix\_w}$	$CLRank_{feat}$	$CLRank_{ins}$
1	AP	OHSUMED	0.284	0.267	0.291	**0.320** (12.7%)	0.257 (−9.5%)
2	WSJ	AP	0.355	0.361	0.370	**0.391** (10.1%)	0.359 (1.1%)
3	WSJ	OHSUMED	0.285	0.271	0.293	**0.309** (8.4%)	0.278 (−2.5%)
4	td2003	td2004	0.178	0.157	0.183	0.185 (3.9%)	**0.190 (6.7%)**
5	hp2003	hp2004	0.647	0.644	0.664	**0.690** (6.6%)	0.655 (1.2%)
6	np2003	np2004	0.504	0.496	0.536	**0.569** (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	$LRank_{std}$	$LRank_{mix}$	$LRank_{mix\_w}$	$CLRank_{feat}$	$CLRank_{ins}$
1	AP	OHSUMED	0.399	0.364	0.417	**0.445** (11.5%)	0.365 (−8.5%)
2	WSJ	AP	0.679	0.689	0.699	**0.726** (6.9%)	0.684 (0.7%)
3	WSJ	OHSUMED	0.418	0.357	0.425	**0.438** (4.8%)	0.408 (−2.4%)
4	td2003	td2004	0.258	0.217	0.266	0.268 (3.9%)	**0.275** (6.6%)
5	hp2003	hp2004	0.679	0.689	0.702	**0.726** (6.9%)	0.684 (0.7%)
6	np2003	np2004	0.532	0.532	0.559	**0.599** (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	-	$\surd$	$\surd$
Feature-based	-	$\surd$	$\surd$
Instance-based	-	$\surd$	**Y**?

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.