Ranking Methods in Machine Learning

A Tutorial Introduction

Shivani Agarwal

Computer Science & Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Example 1: Recommendation Systems
Example 2: Information Retrieval
Example 2: Information Retrieval
Example 2: Information Retrieval

Information - Wikipedia, the free encyclopedia
Information as a concept has many meanings, from everyday usage to technical settings. The concept of information is closely related to notions of...
Etymology - As sensory input - As an influence which leads to...
en.wikipedia.org/wiki/Information - Cached - Similar

Information theory - Wikipedia, the free encyclopedia
Information theory is a branch of applied mathematics and electrical engineering involving the quantification of information. ...
en.wikipedia.org/wiki/Information_theory - Cached - Similar

Information Please
Infoplease.com, a free, authoritative, and respected reference for Internet users, provides a comprehensive encyclopedia, almanac, atlas, dictionary, ...
Countries - United States - This Day In History - Biography
www.infoplease.com/ - Cached - Similar

Local business results for information near Allston, MA - Change location
Federal Reserve Bank: General Information
www.bos.frb.org - (617) 973-3000 - More
Dana-Farber Cancer Institute
www.dana-farber.org - (617) 632-3000 - 95 reviews
Problem: Millions of structures in a chemical library. How do we identify the most promising ones?
Human genetics is now at a critical juncture. The molecular methods used successfully to identify the genes underlying rare mendelian syndromes are failing to find the numerous genes causing more common, familial, non-mendelian diseases...
With the human genome sequence nearing completion, new opportunities are being presented for unravelling the complex genetic basis of nonmendelian disorders based on large-scale genomewide studies . . .

Types of Ranking Problems

Instance Ranking
Label Ranking
Subset Ranking
Rank Aggregation

?
Instance Ranking

doc1 > doc2, 10

doc3 > doc5, 20

...
Label Ranking

doc1
- sports > politics
- health > money
- ...

doc2
- science > sports
- money > politics
- ...

- ...
- ...
Subset Ranking

query 1

doc1 > doc2, doc3 > doc5, ...

query 2

doc2 > doc4, doc11 > doc3, ...

...
Rank Aggregation

query 1
results of search engine 1
results of search engine 2
... desired ranking

query 2
results of search engine 1
results of search engine 2
... desired ranking

...
Types of Ranking Problems

- Instance Ranking
- Label Ranking
- Subset Ranking
- Rank Aggregation

This tutorial...
Tutorial Road Map

Part I: Theory & Algorithms

- Bipartite Ranking
- \( k \)-partite Ranking
- Ranking with Real-Valued Labels
- General Instance Ranking

Part II: Applications

- Applications to Bioinformatics
- Applications to Drug Discovery
- Subset Ranking and Applications to Information Retrieval

Further Reading & Resources
Part I
Theory & Algorithms
[for Instance Ranking]
Bipartite Ranking

Relevant (+)
- doc1+
- doc2+
- doc3+
- ...

Irrelevant (-)
- doc1-
- doc2-
- doc3-
- doc4-
- ...

Bipartite Ranking

- **Instance space** $X$

- **Input:** Training sample $S = (S_+, S_-)$:
  
  $S_+ = (x_1^+, \ldots, x_m^+) \in X^m \quad$ (positive examples)
  
  $S_- = (x_1^-, \ldots, x_n^-) \in X^n \quad$ (negative examples)

- **Output:** Ranking function $f : X \rightarrow \mathbb{R}$

![Diagram showing bipartite graph with positive and negative examples]
Bipartite Ranking

- **Instance space** $X$

- **Input:** Training sample $S = (S_+, S_-)$:
  
  $S_+ = (x_1^+, \ldots, x_m^+) \in X^m$  \hspace{1cm} (positive examples)

  $S_- = (x_1^-, \ldots, x_n^-) \in X^n$  \hspace{1cm} (negative examples)

- **Output:** Ranking function $f : X \rightarrow \mathbb{R}$

- **Expected error:** $\text{er}(f) = P_{(x,x') \sim D_+ \times D_-} [f(x) < f(x')]$

- **Empirical error:** $\widehat{\text{er}}_S(f) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} 1(f(x_i^+) < f(x_j^-))$
Is Bipartite Ranking Different from Binary Classification?

Example 1

\[ f_1 \]

\[ f_2 \]
Is Bipartite Ranking Different from Binary Classification?

Example 1

\[ f_1 \]

\[ f_2 \]

Classification error \( \frac{1}{4} \)
Is Bipartite Ranking Different from Binary Classification?

Example 1

- $f_1$
  - Classification error = $\frac{1}{4}$
  - Ranking error = $\frac{1}{4}$

- $f_2$
  - Classification error = $\frac{1}{4}$
  - Ranking error = $\frac{1}{2}$
Is Bipartite Ranking Different from Binary Classification?

Example 2
Is Bipartite Ranking Different from Binary Classification?

Example 2

Classification error = $\frac{1}{100}$
Is Bipartite Ranking Different from Binary Classification?

Example 2

\[ f \]

Classification error = \( \frac{1}{100} \)

Ranking error = \( \frac{1}{2} \)
Bipartite Ranking: 
Basic Algorithmic Framework

Minimize a convex upper bound on the empirical ranking error, possibly with some regularization, over some class of ranking functions:

$$\min_{f \in F} \left[ \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \ell(f, x_i^+, x_j^-) + \lambda N(f) \right]$$

where

$$\ell(f, x_i^+, x_j^-) : \text{convex upper bound on } 1(f(x_i^+) < f(x_j^-))$$

$$N(f) : \text{regularizer}$$

$$\lambda > 0 : \text{regularization parameter}$$

$$F : \text{class of ranking functions}$$
Bipartite RankSVM Algorithm

\[
\min_{f \in \mathcal{F}_K} \left[ \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \ell_{\text{hinge}}(f, x_i^+, x_j^-) + \frac{\lambda}{2} \|f\|_K^2 \right]
\]

\[
\ell_{\text{hinge}}(f, x_i^+, x_j^-) = \left( 1 - (f(x_i^+) - f(x_j^-)) \right)_+ \quad [u_+ = \max(u, 0)]
\]

\[
\mathcal{F}_K = \text{reproducing kernel Hilbert space (RKHS)}
\]

with kernel function \( K \)

\[
N(f) = \frac{\|f\|_K^2}{2}
\]

[Herbrich et al, 2000; Joachims, 2002; Rakotomamonjy, 2004]
Bipartite RankSVM Algorithm

Introducing slack variables and taking the Lagrangian dual results in the following convex quadratic program (QP) over \( mn \) variables \( \{\alpha_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\} \):

\[
\min_{\alpha} \left[ \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{m} \sum_{l=1}^{n} \alpha_{ij} \alpha_{kl} \phi(x_i^+, x_j^-, x_k^+, x_l^-) - \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_{ij} \right]
\]

subject to \( 0 \leq \alpha_{ij} \leq C \) \( \forall i, j \)

where

\[
\phi(x_i^+, x_j^-, x_k^+, x_l^-) = \left( K(x_i^+, x_k^+) - K(x_i^+, x_l^-) - K(x_j^-, x_k^+) + K(x_j^-, x_l^-) \right)
\]

\[
C = \frac{1}{\lambda mn}
\]

Can be solved using a standard QP solver, or more efficient methods (e.g. Chapelle & Keerthi, 2010).
Bipartite RankBoost Algorithm

\[
\min_{f \in \mathcal{L}(\mathcal{F}_{\text{base}})} \left[ \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \ell_{\exp}(f, x^+_i, x^-_j) \right]
\]

\[
\ell_{\exp}(f, x^+_i, x^-_j) = \exp \left( - \left( f(x^+_i) - f(x^-_j) \right) \right)
\]

\[
\mathcal{L}(\mathcal{F}_{\text{base}}) = \text{linear combinations of functions in some base class } \mathcal{F}_{\text{base}}
\]

[Freund et al, 2003]
Bipartite RankBoost Algorithm

Input: \((S_+, S_-) \in X^m \times X^n\).

Initialize: \(D_1(x_i^+, x_j^-) = \frac{1}{mn}\) for all \(i \in \{1, \ldots, m\}, j \in \{1, \ldots, n\}\).

For \(t = 1, \ldots, T\):

- Train weak learner using distribution \(D_t\); get weak ranker \(f_t \in \mathcal{F}_{\text{base}}\).
- Choose \(\alpha_t \in \mathbb{R}\).
- Update: \(D_{t+1}(x_i^+, x_j^-) = \frac{1}{Z_t} D_t(x_i^+, x_j^-) \exp\left(-\alpha_t \left(f_t(x_i^+) - f_t(x_j^-)\right)\right)\)

where \(Z_t = \sum_{i=1}^{m} \sum_{j=1}^{n} D_t(x_i^+, x_j^-) \exp\left(-\alpha_t \left(f_t(x_i^+) - f_t(x_j^-)\right)\right)\).

Output final ranking: \(f(x) = \sum_{t=1}^{T} \alpha_t f_t(x)\).
Bipartite RankNet Algorithm

$$\min_{f \in \mathcal{F}_{\text{neural}}} \left[ \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \ell_{\text{logistic}}(f, x_i^+, x_j^-) \right]$$

$$\ell_{\text{logistic}}(f, x_i^+, x_j^-) = \log \left( 1 + \exp \left( - (f(x_i^+) - f(x_j^-)) \right) \right)$$

$$\mathcal{F}_{\text{neural}} = \text{functions represented by some class of neural networks}$$

[Burges et al, 2005]
$k$-partite Ranking

Rating $k$

\[ \text{doc}^k_1 \quad \text{doc}^k_2 \quad \text{doc}^k_3 \ldots \]

\vdots

Rating 2

\[ \text{doc}^2_1 \quad \text{doc}^2_2 \quad \text{doc}^2_3 \quad \text{doc}^2_4 \ldots \]

Rating 1

\[ \text{doc}^1_1 \quad \text{doc}^1_2 \quad \text{doc}^1_3 \quad \text{doc}^1_4 \ldots \]
**$k$-partite Ranking**

- **Instance space** $X$

- **Input:** Training sample $S = (S_1, S_2, \ldots, S_k)$:
  
  $S_k = (x^k_1, \ldots, x^k_{n_k}) \in X^{n_k}$  \hspace{1cm} (examples of rating $k$)
  
  $S_2 = (x^2_1, \ldots, x^2_{n_2}) \in X^{n_2}$  \hspace{1cm} (examples of rating $2$)
  
  $S_1 = (x^1_1, \ldots, x^1_{n_1}) \in X^{n_1}$  \hspace{1cm} (examples of rating $1$)

- **Output:** Ranking function $f : X \rightarrow \mathbb{R}$

- **Empirical error:**

  \[
  \hat{\text{er}}_S(f) = \left( \frac{1}{\sum_{1 \leq a < b \leq k} n_a n_b} \right) \sum_{1 \leq a < b \leq k} \sum_{i=1}^{n_b} \sum_{j=1}^{n_a} (b - a) \, 1(f(x^b_i) < f(x^a_j))
  \]


**k-partite Ranking: Basic Algorithmic Framework**

Minimize a convex upper bound on the empirical ranking error, possibly with some regularization, over some class of ranking functions:

\[
\min_{f \in \mathcal{F}} \left[ \left( \sum_{1 \leq a < b \leq k} \frac{1}{n_a n_b} \right) \sum_{1 \leq a < b \leq k} \sum_{i=1}^{n_b} \sum_{j=1}^{n_a} \ell(f, x_i^b, x_j^a, (b - a)) + \lambda N(f) \right]
\]

where

\[
\ell(f, x_i^b, x_j^a, (b - a)) : \text{convex upper bound on } (b - a) \mathbbm{1}(f(x_i^b) < f(x_j^a))
\]

\[
N(f) : \text{regularizer}
\]

\[
\lambda > 0 : \text{regularization parameter}
\]

\[
\mathcal{F} : \text{class of ranking functions}
\]
Ranking with Real-Valued Labels

\[ \text{doc1} \quad y_1 \]

\[ \text{doc2} \quad y_2 \]

\[ \text{doc3} \quad y_3 \]

\[ \ldots \]
Ranking with Real-Valued Labels

- Instance space $X$
- Real-valued labels $Y = \mathbb{R}$
- **Input:** Training sample $S = ((x_1, y_1), \ldots, (x_m, y_m)) \in (X \times \mathbb{R})^m$
- **Output:** Ranking function $f : X \rightarrow \mathbb{R}$

- Empirical error:

$$\widehat{\text{er}}_S(f) = \frac{1}{\binom{m}{2}} \sum_{1 \leq i < j \leq m} |y_i - y_j| 1 \left( (y_i - y_j)(f(x_i) - f(x_j)) < 0 \right)$$
Ranking with Real-Valued Labels: Basic Algorithmic Framework

Minimize a convex upper bound on the empirical ranking error, possibly with some regularization, over some class of ranking functions:

$$\min_{f \in \mathcal{F}} \left[ \frac{1}{\binom{m}{2}} \sum_{1 \leq i < j \leq m} \ell(f, (x_i, y_i), (x_j, y_j)) + \lambda N(f) \right]$$

where

$$\ell(f, (x_i, y_i), (x_j, y_j)) : \text{convex upper bound on}$$

$$|y_i - y_j| \mathbf{1} \left( (y_i - y_j)(f(x_i) - f(x_j)) < 0 \right)$$

$$N(f) : \text{regularizer}$$

$$\lambda > 0 : \text{regularization parameter}$$

$$\mathcal{F} : \text{class of ranking functions}$$
General Instance Ranking

doc1 > doc1', r1

doc2 > doc2', r2

...
General Instance Ranking

- Instance space $X$

- **Input:** Training sample $S = ((x_1, x'_1, r_1), \ldots, (x_m, x'_m, r_m)) \in (X^2 \times \mathbb{R}_+)^m$

- **Output:** Ranking function $f : X \rightarrow \mathbb{R}$
General Instance Ranking

- Instance space $X$
- **Input:** Training sample $S = ((x_1, x'_1, r_1), \ldots, (x_m, x'_m, r_m)) \in (X^2 \times \mathbb{R}_+)^m$
- **Output:** Ranking function $f : X \rightarrow \mathbb{R}$

- Empirical error: $\widehat{er}_S(f) = \frac{1}{m} \sum_{i=1}^{m} r_i 1(f(x_i) < f(x'_i))$
General Instance Ranking: Basic Algorithmic Framework

Minimize a convex upper bound on the empirical ranking error, possibly with some regularization, over some class of ranking functions:

$$\min_{f \in \mathcal{F}} \left[ \frac{1}{m} \sum_{i=1}^{m} \ell(f, x_i, x'_i, r_i) + \lambda N(f) \right]$$

where

$$\ell(f, x_i, x'_i, r_i) : \text{convex upper bound on } r_i 1(f(x_i) < f(x'_i))$$

$$N(f) : \text{regularizer}$$

$$\lambda > 0 : \text{regularization parameter}$$

$$\mathcal{F} : \text{class of ranking functions}$$
General RankSVM Algorithm

\[
\min_{f \in \mathcal{F}_K} \left[ \frac{1}{m} \sum_{i=1}^{m} \ell_{\text{hinge}}(f, x_i, x'_i, r_i) + \frac{\lambda}{2} \|f\|_K^2 \right]
\]

\[
\ell_{\text{hinge}}(f, x_i, x'_i, r_i) = \left( r_i - (f(x_i) - f(x'_i)) \right)_+ \quad \left[ u_+ = \max(u, 0) \right]
\]

\[
\mathcal{F}_K = \text{reproducing kernel Hilbert space (RKHS)}
\]

\[
\text{with kernel function } K
\]

\[
N(f) = \frac{\|f\|_K^2}{2}
\]

[Herbrich et al, 2000; Joachims, 2002]
General RankBoost Algorithm

$$\min_{f \in \mathcal{L}(\mathcal{F}_{\text{base}})} \left[ \frac{1}{m} \sum_{i=1}^{m} \ell_{\exp}(f, x_i, x'_i, r_i) \right]$$

$$\ell_{\exp}(f, x_i, x'_i, r_i) = r_i \exp \left( - \left( f(x_i) - f(x'_i) \right) \right)$$

$$\mathcal{L}(\mathcal{F}_{\text{base}}) = \text{linear combinations of functions in some base class } \mathcal{F}_{\text{base}}$$

[Freund et al, 2003]
General RankNet Algorithm

$$\min_{f \in \mathcal{F}_{\text{neural}}} \left[ \frac{1}{m} \sum_{i=1}^{m} \ell_{\text{logistic}}(f, x_i, x'_i, r_i) \right]$$

$$\ell_{\text{logistic}}(f, x_i, x'_i, r_i) = r_i \log \left( 1 + \exp \left( - (f(x_i) - f(x'_i)) \right) \right)$$

$$\mathcal{F}_{\text{neural}} = \text{functions represented by some class of neural networks}$$

[Burges et al, 2005]
Tutorial Road Map

Part I: Theory & Algorithms
- Bipartite Ranking
- \(k\)-partite Ranking
- Ranking with Real-Valued Labels
- General Instance Ranking
- RankSVM
- RankBoost
- RankNet

Part II: Applications
- Applications to Bioinformatics
- Applications to Drug Discovery
- Subset Ranking and Applications to Information Retrieval

Further Reading & Resources
Part II
Applications
[and Subset Ranking]
Human genetics is now at a critical juncture. The molecular methods used successfully to identify the genes underlying rare mendelian syndromes are failing to find the numerous genes causing more common, familial, non-mendelian diseases . . .
With the human genome sequence nearing completion, new opportunities are being presented for unravelling the complex genetic basis of nonmendelian disorders based on large-scale genomewide studies . . .
Identifying Genes Relevant to a Disease Using Microarray Gene Expression Data

Biological samples \((d)\)

Genes \((N)\)

\(N \gg d\)
Identifying Genes Relevant to a Disease Using Microarray Gene Expression Data

\[
\text{Genes } (N) \quad \text{Biological samples } (d)
\]

\[N \gg d\]
### Identifying Genes Relevant to a Disease Using Microarray Gene Expression Data

<table>
<thead>
<tr>
<th>Biological samples ($d$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genes ($N$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N \gg d$</td>
</tr>
</tbody>
</table>
Identifying Genes Relevant to a Disease Using Microarray Gene Expression Data

Biological samples \( (d) \)

Genes \( (N) \)

\( N \gg d \)
Identifying Genes Relevant to a Disease Using Microarray Gene Expression Data

\[ N \gg d \]
Identifying Genes Relevant to a Disease Using Microarray Gene Expression Data

Biological samples \( (d) \)

Genes \( (N) \)

\( N \gg d \)
Formulation as a Bipartite Ranking Problem

Relevant

Not relevant
# Microarray Gene Expression Data Sets

[Golub et al, 1999; Alon et al, 1999]

<table>
<thead>
<tr>
<th>Data Set</th>
<th>No. of Genes</th>
<th>No. of Tissue Samples</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leukemia</td>
<td>7129</td>
<td>72</td>
<td>25 AML / 47 ALL</td>
</tr>
<tr>
<td>Colon cancer</td>
<td>2000</td>
<td>62</td>
<td>40 tumor / 22 normal</td>
</tr>
</tbody>
</table>
# Selection of Training Genes

## Leukemia

<table>
<thead>
<tr>
<th>Positive genes: Markers for AML/ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myeloperoxidase</td>
</tr>
<tr>
<td>CD13</td>
</tr>
<tr>
<td>CD33</td>
</tr>
<tr>
<td>HOXA9 Homeo box A9</td>
</tr>
<tr>
<td>V-myb</td>
</tr>
<tr>
<td>CD19</td>
</tr>
<tr>
<td>CD10 (CALLA)</td>
</tr>
<tr>
<td>TCL1 (T cell leukemia)</td>
</tr>
<tr>
<td>C-myb</td>
</tr>
<tr>
<td>Deoxyhypusine synthase</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>157 genes involved in physiological cellular functions</td>
</tr>
</tbody>
</table>

## Colon cancer

<table>
<thead>
<tr>
<th>Positive genes: Markers for colon cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phospholipase A2</td>
</tr>
<tr>
<td>Keratin 6 isoform</td>
</tr>
<tr>
<td>PTP-H1</td>
</tr>
<tr>
<td>TF-III A</td>
</tr>
<tr>
<td>V-raf oncogene</td>
</tr>
<tr>
<td>MAPK kinase 1</td>
</tr>
<tr>
<td>CEA</td>
</tr>
<tr>
<td>Oncoprotein 18</td>
</tr>
<tr>
<td>PEP carboxykinase</td>
</tr>
<tr>
<td>ERK kinase 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Negative genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>56 genes involved in physiological cellular functions</td>
</tr>
</tbody>
</table>
Top-Ranking Genes for Leukemia Returned by RankBoost

- ♦ Known marker; ♦ Potential marker;
- ■ Known therapeutic target; ■ Potential therapeutic target;
- x No link found.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Relevance Summary</th>
<th>t-Statistic Rank</th>
<th>Pearson Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. KIAA0220</td>
<td>■</td>
<td>6628</td>
<td>2461</td>
</tr>
<tr>
<td>2. G-gamma globin</td>
<td>♦</td>
<td>3578</td>
<td>3567</td>
</tr>
<tr>
<td>3. Delta-globin</td>
<td>♦</td>
<td>3663</td>
<td>3532</td>
</tr>
<tr>
<td>4. Brain-expressed HHCPA78 homolog</td>
<td>■</td>
<td>6734</td>
<td>2390</td>
</tr>
<tr>
<td>5. Myeloperoxidase</td>
<td>♦</td>
<td>139</td>
<td>6573</td>
</tr>
<tr>
<td>6. Disulfide isomerase precursor</td>
<td>■</td>
<td>6650</td>
<td>575</td>
</tr>
<tr>
<td>7. Nucleophosmin</td>
<td>♦</td>
<td>405</td>
<td>1115</td>
</tr>
<tr>
<td>8. CD34</td>
<td>♦</td>
<td>6732</td>
<td>643</td>
</tr>
<tr>
<td>9. Elongation factor-1/β</td>
<td>x</td>
<td>4460</td>
<td>3413</td>
</tr>
<tr>
<td>10. CD24</td>
<td>♦</td>
<td>81</td>
<td>1</td>
</tr>
<tr>
<td>11. 60S ribosomal protein L23</td>
<td>■</td>
<td>1950</td>
<td>73</td>
</tr>
<tr>
<td>12. 5-aminolevulinic acid synthase</td>
<td>■</td>
<td>4750</td>
<td>3351</td>
</tr>
</tbody>
</table>

[Agarwal & Sengupta, 2009]
Biological Validation

KIAA0220

[Agarwal et al, 2010]
Problem: Millions of structures in a chemical library. How do we identify the most promising ones?
Formulation as a Ranking Problem with Real-Valued Labels

\[ pIC_{50} = 5.6718 \]

\[ pIC_{50} = 8.2991 \]

\[ pIC_{50} = 4.1317 \]

\[ \ldots \]
# Cheminformatics Data Sets

[Sutherland et al, 2004]

<table>
<thead>
<tr>
<th>Data Set</th>
<th>No. of Compounds</th>
<th>No. of Chemical (2.5D) Descriptors</th>
<th>pIC$_{50}$ Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>DHFR inhibitors</td>
<td>361</td>
<td>70</td>
<td>3.3 – 9.8</td>
</tr>
<tr>
<td>COX2 inhibitors</td>
<td>292</td>
<td>74</td>
<td>4.0 – 9.0</td>
</tr>
</tbody>
</table>
### DHFR Results Using RankSVM

**2.5D chemical descriptors**  
Gaussian kernel  

<table>
<thead>
<tr>
<th>Training size</th>
<th>SVR</th>
<th>RankSVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>0.4755</td>
<td>0.4601</td>
</tr>
<tr>
<td>48</td>
<td>0.3430</td>
<td>0.3509</td>
</tr>
<tr>
<td>72</td>
<td>0.2840</td>
<td>0.2726</td>
</tr>
<tr>
<td>96</td>
<td>0.2483</td>
<td>0.2351</td>
</tr>
<tr>
<td>120</td>
<td>0.2171</td>
<td>0.2121</td>
</tr>
<tr>
<td>144</td>
<td><strong>0.2023</strong></td>
<td>0.2032</td>
</tr>
<tr>
<td>168</td>
<td>0.2019</td>
<td><strong>0.1817</strong></td>
</tr>
<tr>
<td>192</td>
<td>0.1808</td>
<td>0.1749</td>
</tr>
<tr>
<td>216</td>
<td>0.1816</td>
<td>0.1722</td>
</tr>
<tr>
<td>237</td>
<td>0.1714</td>
<td>0.1681</td>
</tr>
</tbody>
</table>

**FP2 molecular fingerprints**  
Tanimoto kernel  

<table>
<thead>
<tr>
<th>Training size</th>
<th>SVR</th>
<th>RankSVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>0.3793</td>
<td>0.3546</td>
</tr>
<tr>
<td>48</td>
<td>0.2905</td>
<td>0.2896</td>
</tr>
<tr>
<td>72</td>
<td>0.2517</td>
<td>0.2421</td>
</tr>
<tr>
<td>96</td>
<td>0.2343</td>
<td>0.2201</td>
</tr>
<tr>
<td>120</td>
<td>0.2147</td>
<td>0.2052</td>
</tr>
<tr>
<td>144</td>
<td>0.2166</td>
<td>0.1988</td>
</tr>
<tr>
<td>168</td>
<td>0.2096</td>
<td>0.1966</td>
</tr>
<tr>
<td>192</td>
<td>0.2056</td>
<td>0.1962</td>
</tr>
<tr>
<td>216</td>
<td>0.1907</td>
<td>0.1787</td>
</tr>
<tr>
<td>237</td>
<td>0.1924</td>
<td>0.1798</td>
</tr>
</tbody>
</table>

[Agarwal et al, 2010]
Application to Information Retrieval (IR)
Learning to Rank in IR

Google™

query 1

Google Search  I'm Feeling Lucky

q1
Learning to Rank in IR
Learning to Rank in IR

q1
rel1
Learning to Rank in IR
Learning to Rank in IR
Learning to Rank in IR

q1
rel1 q2
rel2
Learning to Rank in IR
Learning to Rank in IR
Learning to Rank in IR

q1
rel1  q2
rel2  q3
rel3
Learning to Rank in IR
General Subset Ranking

query 1

doc1 > doc2, doc3 > doc5, ...

query 2

doc2 > doc4, doc11 > doc3, ...

...
General Subset Ranking

- Query space $Q$
- Document space $D$
- Query-document feature mapping $\phi : Q \times D \to \mathbb{R}^d$
- Input: Training sample $S = (S^1, \ldots, S^m)$:
  
  $$S^i = ((\phi_1^i, \phi_1^{i'}), \ldots, (\phi_{n_i}^i, \phi_{n_i}^{i'})) \in (\mathbb{R}^d \times \mathbb{R}^d)^{n_i}$$

  where
  
  $$\phi_j^i = \phi(q^i, d_j^i), \quad \phi_j^{i'} = \phi(q^i, d_j^{i'})$$

- Output: Ranking function $f : \mathbb{R}^d \to \mathbb{R}$
Subset Ranking with Real-Valued Relevance Labels
Subset Ranking with Real-Valued Relevance Labels

- Query space $Q$
- Document space $D$
- Query-document feature mapping $\phi : Q \times D \rightarrow \mathbb{R}^d$
- Input: Training sample $S = (S^1, \ldots, S^m)$:
  
  $$S^i = \left((\phi_1^i, y_1^i), \ldots, (\phi_{n_i}^i, y_{n_i}^i)\right) \in (\mathbb{R}^d \times \mathbb{R})^{n_i}$$

  where
  
  $$\phi_j^i = \phi(q^i, d_j^i), \quad y_j^i = \text{relevance of } d_j^i \text{ to } q^i$$

- Output: Ranking function $f : \mathbb{R}^d \rightarrow \mathbb{R}$
RankSVM Applied to IR/Subset Ranking

**Standard RankSVM**

\[
\min_{f \in \mathcal{F}_K} \left[ \left( \frac{1}{\sum_{i=1}^{m} \binom{n_i}{2}} \right) \sum_{i=1}^{m} \sum_{1 \leq j < k \leq n_i} \ell_{\text{hinge}} \left( f, (\phi_j^i, y_j^i), (\phi_k^i, y_k^i) \right) + \frac{\lambda}{2} \| f \|_K^2 \right]
\]

\[
\ell_{\text{hinge}} \left( f, (\phi_j^i, y_j^i), (\phi_k^i, y_k^i) \right) = \left( 1 - \left( \text{sign}(y_j^i - y_k^i) \cdot (f(\phi_j^i) - f(\phi_k^i)) \right) \right)_+ ,
\]

convex upper bound on

\[
1 \left( (y_j^i - y_k^i)(f(\phi_j^i) - f(\phi_k^i)) < 0 \right)
\]

[Joachims, 2002]
RankSVM Applied to IR/Subset Ranking

RankSVM with Query Normalization & Relevance Weighting

\[
\min_{f \in \mathcal{F}_K} \left[ \frac{1}{m} \sum_{i=1}^{m} \left[ \frac{1}{n_i^2} \sum_{1 \leq j < k \leq n_i} \ell_{hinge}^{rel} (f, (\phi^i_j, y^i_j), (\phi^i_k, y^i_k)) + \frac{\lambda}{2} \| f \|_2^2 \right] \right]
\]

\[
\ell_{hinge}^{rel} (f, (\phi^i_j, y^i_j), (\phi^i_k, y^i_k)) = \left( \| y^i_j - y^i_k \| - \left( \text{sign} (y^i_j - y^i_k) \cdot (f(\phi^i_j) - f(\phi^i_k)) \right) \right)_+,
\]

convex upper bound on

\[
\| y^i_j - y^i_k \| 1 \left( (y^i_j - y^i_k)(f(\phi^i_j) - f(\phi^i_k)) < 0 \right)
\]

[Agarwal & Collins, 2010; also Cao et al, 2006]
Ranking Performance Measures in IR

Mean Average Precision (MAP)

Binary Labels: \( y_j \in \{0, 1\} \)

\[
\text{MAP}_S(f) = \frac{1}{m} \sum_{i=1}^{m} \left[ \frac{1}{|\{j : y_j^i = 1\}|} \sum_{j: y_j^i = 1} \text{prec}_{r_j^i}(f) \right]
\]

\( r_j^i = \) rank of document \( d_j^i \) for query \( q^i \)

\( \text{prec}_{r}^i(f) = \) fraction of positives in top \( r \) documents for query \( q^i \)
Normalized Discounted Cumulative Gain (NDCG)

General Real-Valued Labels: $y_j \in \mathbb{R}$

$$NDCG_S(f) = \frac{1}{m} \sum_{i=1}^{m} \left[ \frac{1}{Z_i} \sum_{r=1}^{n_i} \frac{2^{y_i^r} - 1}{\log_2(r + 1)} \right]$$

$\pi_r^i = \text{index of document ranked at position } r \text{ for query } q^i$

$Z_i = \text{normalization constant}$

$$NDCG@k_S(f) = \frac{1}{m} \sum_{i=1}^{m} \left[ \frac{1}{Z_i} \sum_{r=1}^{k} \frac{2^{y_i^r} - 1}{\log_2(r + 1)} \right]$$
Ranking Algorithms for Optimizing MAP/NDCG

- SVMMAP [Yue et al. 2007]
- SVMNDCG [Chapelle et al. 2007]
- LambdaRank [Burges et al. 2007]
- AdaRank [Xu & Li 2007]
- Regression-based algorithm [Cossock & Zhang 2008]
- SoftRank [Taylor et al. 2008]
- SmoothRank [Chapelle & Wu 2010]
## LETOR 3.0/OHSUMED Data Set

[Liu et al, 2007]

<table>
<thead>
<tr>
<th>No. of Queries</th>
<th>Relevance Labels</th>
<th>Total no. of Query-Doc Pairs</th>
<th>Avg. no. of Docs/Query</th>
<th>No. of Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>106</td>
<td>2 : definitely relevant</td>
<td>16,140</td>
<td>152</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>1 : partially relevant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 : not relevant</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
OHSUMED Results – NDCG

- RankSVM [incomplete]
- RankSVM-Primal
- RankSVM-Struct
- SmoothRank
- AdaRank-MAP
- AdaRank-NDCG
- SVMMAP
- RankSVM with QN & RW
Further Reading & Resources

[Incomplete!]
Early Papers on Ranking


Generalization Bounds for Ranking


S. Agarwal and S. Sengupta, Ranking genes by relevance to a disease, CSB 2009.

Other Applications

Natural Language Processing

Collaborative Filtering

Manhole Event Prediction
IR Ranking Algorithms

Y. Cao, J. Xu, T.-Y. Liu, H. Li, Y. Hunag, and H.W. Hon, Adapting ranking SVM to document retrieval, SIGIR 2006.


IR Ranking Algorithms


S. Agarwal and M. Collins, Maximum margin ranking algorithms for information retrieval, ECIR 2010.
NIPS Workshop 2005
Learning to Rank

SIGIR Workshops 2007-2009
Learning to Rank for Information Retrieval

NIPS Workshop 2009
Advances in Ranking

American Institute of Mathematics
Workshop in Summer 2010
The Mathematics of Ranking
Tutorial Articles & Books

