Probabilistic Modeling on Rankings

Jose A. Lozano   Ekhine Irurozki

Intelligent Systems Group
University of the Basque Country UPV/EHU

ACML, Singapore, November 4th, 2012

Outline of the talk

Problems we are interested in

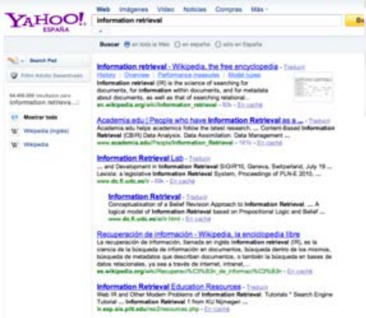
Identity-tracking

Problems we are interested in

Preferences

Problems we are interested in

Information retrieval



UPV - EHU

Probabilistic Modeling on Ranking

Remarks

- Social science
- Machine learning: ECML, ICML, UAI, NIPS, JMLR
- NOT “learning to rank”
- A model can be explained from many different points of view
- *Probabilistic graphical model bias*

Permutations

What are permutations?

- A permutation σ can be seen as a bijection between the set $\{1, 2, \ldots, n\}$ onto itself:

$$\begin{array}{rcl} \sigma : \{1, 2, \ldots, n\} & \longrightarrow & \{1, 2, \ldots, n\} \\ i & \longmapsto & \sigma(i) \end{array}$$

- Interpretation
  - $\sigma(i)$ represents the rank associated to the i-th element
  - $\sigma^{-1}(i)$ represents the i-th ranked element
- S_n is the set of permutations of n elements, then (S_n, o) forms the symmetric group:

$$\sigma_1 o\, \sigma_2(i) = \sigma_1(\sigma_2(i))$$

Permutations

Notation and representation

- The identity permutation $(1\ 2 \ldots\ n)$ will be denoted by e
- A permutation σ can be equivalently represented as a permutation matrix $M = [m_{ij}]$ where:

$$m_{ij} = \begin{cases} 1 & \text{if } \sigma(i) = j \\ 0 & \text{otherwise} \end{cases}$$

Learning probability distributions over permutations

(1 2 3 4 5 6 7)
(2 3 1 7 6 5 4)
(6 7 1 3 4 2 5)
(5 2 7 6 3 4 1)
(5 1 7 2 4 3 6)
(3 5 7 1 2 4 6)
$\vdots$

$$\begin{aligned} p : S_n &\longrightarrow [0,1] \\ \sigma &\longmapsto p(\sigma) \end{aligned}$$

Learning probability distributions over permutations

$$
\begin{array}{l}
(1\,2\,3\,4\,5\,6\,7 - - - - \ldots)\\
(2\,3\,1\,7\,6\,5\,4 - - - - \ldots)\\
(6\,7\,1\,3\,4\,2\,5 - - - - \ldots)\\
(5\,2\,7\,6\,3\,4\,1 - - - - \ldots)\\
(5\,1\,7\,2\,4\,3\,6 - - - - \ldots)\\
(3\,5\,7\,1\,2\,4\,6 - - - - \ldots)\\
\vdots
\end{array}
$$

$$
\begin{aligned}
p : S_n &\longrightarrow [0,1]\\
\sigma &\longmapsto p(\sigma)
\end{aligned}
$$

UPV - EHU

Learning probability distributions over permutations

$1 \succ 2 \succ 4,\ 5$

$2,\ 3,\ 1 \succ 7 \succ 6,\ 5,\ 4$

$6,\ 7,\ 1,\ 3 \succ 4 \succ 2$

$5,\ 2 \succ 7,\ 6 \succ 3,\ 4 \succ 1$

$\vdots$

$$\begin{aligned} p : S_n &\longrightarrow [0,1] \\ \sigma &\longmapsto p(\sigma) \end{aligned}$$

Learning probability distributions over permutations

$$
\begin{array}{l}
(1\ 2\ 3\ 4\ -\ -\ \ldots -\ -\ 5\ 6\ 7)\\
(2\ 3\ 1\ 7\ -\ -\ \ldots -\ -\ 6\ 5\ 4)\\
(6\ 7\ 1\ 3\ -\ -\ \ldots -\ -\ 4\ 2\ 5)\\
(5\ 2\ 7\ 6\ -\ -\ \ldots -\ -\ 3\ 4\ 1)\\
(5\ 1\ 7\ 2\ -\ -\ \ldots -\ -\ 4\ 3\ 6)\\
(3\ 5\ 7\ 1\ -\ -\ \ldots -\ -\ 2\ 4\ 6)\\
\vdots
\end{array}
$$

$$
\begin{array}{rcl}
p : S_n & \longrightarrow & [0,1]\\
\sigma & \longmapsto & p(\sigma)
\end{array}
$$

UPV - EHU

Learning probability distributions over permutations

$1 \succ 5$

$1 \succ 7$

$3 \succ 4$

$5 \succ 7$

$\vdots$

$$\begin{array}{rcl} p : S_n & \longrightarrow & [0,1] \\ \sigma & \longmapsto & p(\sigma) \end{array}$$

Representing probability distributions over permutation

Problems

- How many parameters are needed?

$$
\begin{array}{ccc}
p(1\;2\;3\;4\;5) & , & p(2\;1\;3\;4\;5) \\
p(3\;2\;1\;4\;5) & , & \dots \\
\dots & , & p(5\;4\;3\;2\;1)
\end{array}
$$

- $5! - 1$ parameters. In general $n! - 1$

Representing probability distributions over permutation

Problems

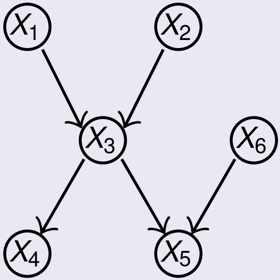
- How many parameters are needed?

$$
\begin{array}{ccc}
p(1\ 2\ 3\ 4\ 5) & , & p(2\ 1\ 3\ 4\ 5) \\
p(3\ 2\ 1\ 4\ 5) & , & \dots \\
\dots & , & p(5\ 4\ 3\ 2\ 1)
\end{array}
$$

- $5! - 1$ parameters. In general $n! - 1$

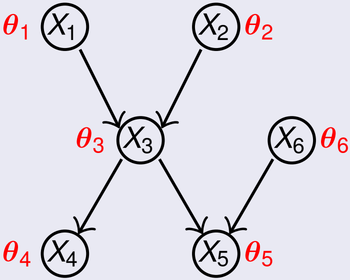
UPV - EHU

Bayesian networks



$$p(\mathbf{x}) = p(x_1) \cdot p(x_2) \cdot p(x_3|x_1, x_2) \cdot p(x_4|x_3) \cdot p(x_5|x_3, x_6) \cdot p(x_6)$$

Bayesian networks

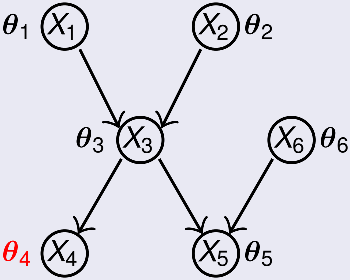


$$p(\mathbf{x}) = p(x_1) \cdot p(x_2) \cdot p(x_3|x_1, x_2) \cdot p(x_4|x_3) \cdot p(x_5|x_3, x_6) \cdot p(x_6)$$

- EHU

Bayesian networks

$$\theta_4 = (\theta_{41}, \theta_{42})$$

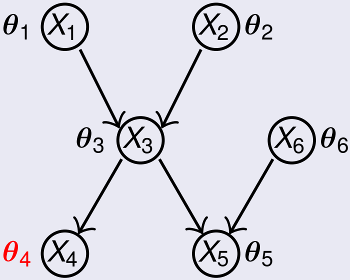


$$p(\mathbf{x}) = p(x_1) \cdot p(x_2) \cdot p(x_3|x_1, x_2) \cdot p(x_4|x_3) \cdot p(x_5|x_3, x_6) \cdot p(x_6)$$

Bayesian networks

$$\boldsymbol{\theta}_4 = (\boldsymbol{\theta}_{41}, \boldsymbol{\theta}_{42})$$

$$\begin{cases} \theta_{411} = p(X_4 = 0 \mid X_3 = 0) \\ \theta_{412} = p(X_4 = 1 \mid X_3 = 0) \end{cases}$$

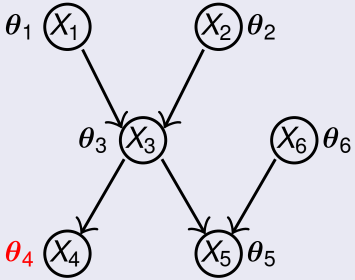


$$p(\mathbf{x}) = p(x_1) \cdot p(x_2) \cdot p(x_3|x_1, x_2) \cdot p(x_4|x_3) \cdot p(x_5|x_3, x_6) \cdot p(x_6)$$

Bayesian networks

$$\theta_4 = (\theta_{41}, \theta_{42})$$

$$\begin{cases} \theta_{411} = p(X_4 = 0 \mid X_3 = 0) \\ \theta_{412} = p(X_4 = 1 \mid X_3 = 0) \end{cases}$$



General case

$$\theta_{ijk} = p(X_i = x_i^k | \boldsymbol{Pa}_i = \boldsymbol{pa}_i^j)$$

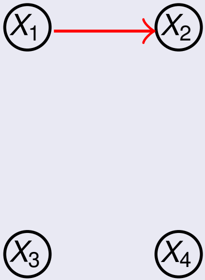
$$p(\mathbf{x}) = p(x_1) \cdot p(x_2) \cdot p(x_3|x_1, x_2) \cdot p(x_4|x_3) \cdot p(x_5|x_3, x_6) \cdot p(x_6)$$

Bayesian networks for permutations

$$X_1 = 1 \Rightarrow X_2 \neq 1$$

Bayesian networks for permutations

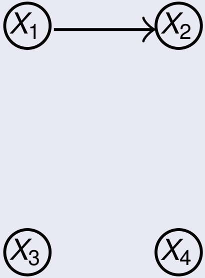
$$X_1 = 1 \Rightarrow X_2 \neq 1$$



Bayesian networks for permutations

$$X_1 = 1 \Rightarrow X_2 \neq 1$$

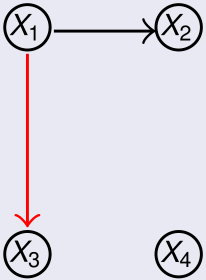
$$X_1 = 1 \Rightarrow X_3 \neq 1$$



Bayesian networks for permutations

$$X_1 = 1 \Rightarrow X_2 \neq 1$$

$$X_1 = 1 \Rightarrow X_3 \neq 1$$

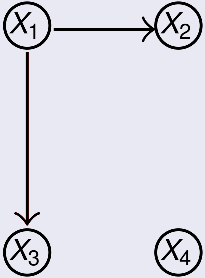


Bayesian networks for permutations

$$X_1 = 1 \Rightarrow X_2 \neq 1$$

$$X_1 = 1 \Rightarrow X_3 \neq 1$$

$$X_1 = 1 \Rightarrow X_4 \neq 1$$

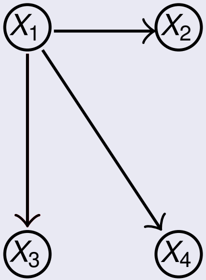


Bayesian networks for permutations

$$X_1 = 1 \Rightarrow X_2 \neq 1$$

$$X_1 = 1 \Rightarrow X_3 \neq 1$$

$$X_1 = 1 \Rightarrow X_4 \neq 1$$

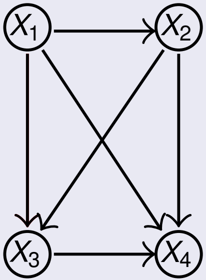


Bayesian networks for permutations

$$X_1 = 1 \Rightarrow X_2 \neq 1$$

$$X_1 = 1 \Rightarrow X_3 \neq 1$$

$$X_1 = 1 \Rightarrow X_4 \neq 1$$

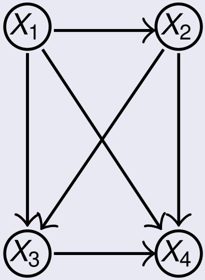


Bayesian networks for permutations

$$X_1 = 1 \Rightarrow X_2 \neq 1$$

$$X_1 = 1 \Rightarrow X_3 \neq 1$$

$$X_1 = 1 \Rightarrow X_4 \neq 1$$



$$p(\mathbf{x}) = p(x_1) \cdot p(x_2|x_1) \cdot p(x_3|x_1, x_2) \cdot p(x_4|x_1, x_2, x_3)$$

Inference over permutations

Recommender systems (Sun and Lebanon, 2012)

- Which is the most preferred film given the current personal rankings?

$$\arg \max_{i \neq 3,7,9,15,4} p(\sigma^{-1}(1) = i \,|3 \succ 7,\ 9,\ 15 \succ 4)$$

- Which is the most probable order of the films given the current rankings?

$$\arg \max_{\sigma} p(\sigma \,|3 \succ 7,\ 9,\ 15 \succ 4)$$

Inference over permutations

Information retrieval

- Label ranking (Cheng and Hüllermeier, 2009; Chen et al., 2010): finding the ranking that maximizes the probability

Outline of the talk

UPV - EHU

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

B1

B2

B3

B4

B5

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

B1 $\longrightarrow$

B2

B3

B4

B5

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

B1 $\longrightarrow$ 5.6

B2

B3

B4

B5

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$$
\begin{array}{lcl}
B1 & \longrightarrow & 5.6 \\
B2 & \longrightarrow & 4.9 \\
B3 & & \\
B4 & & \\
B5 & &
\end{array}
$$

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$$\text{B1} \longrightarrow 5.6$$

$$\text{B2} \longrightarrow 4.9$$

$$\text{B3} \longrightarrow 9.3$$

$$\text{B4} \longrightarrow 2.1$$

$$\text{B5} \longrightarrow 3.4$$

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$$B1 \longrightarrow 5.6$$

$$B2 \longrightarrow 4.9$$

$$B3 \longrightarrow 9.3 \qquad (\_,\_,\_,\_,\_)$$

$$B4 \longrightarrow 2.1$$

$$B5 \longrightarrow 3.4$$

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$B1 \longrightarrow 5.6$

$B2 \longrightarrow 4.9$

$B3 \longrightarrow 9.3$ $\qquad (\_,\_,\_,\_,\_)$

$B4 \longrightarrow 2.1$

$B5 \longrightarrow 3.4$

UPV - EHU

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$$B1 \longrightarrow 5.6$$

$$B2 \longrightarrow 4.9$$

$$B3 \longrightarrow 9.3 \qquad (\_, \_, \_, 1, \_)$$

$$B4 \longrightarrow 2.1$$

$$B5 \longrightarrow 3.4$$

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$$B1 \longrightarrow 5.6$$

$$B2 \longrightarrow 4.9$$

$$B3 \longrightarrow 9.3 \qquad (\_, \_, \_, 1, \_)$$

$$B4 \longrightarrow 2.1$$

$$B5 \longrightarrow 3.4$$

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$$B1 \longrightarrow 5.6$$

$$B2 \longrightarrow 4.9$$

$$B3 \longrightarrow 9.3 \qquad (\_, \_, \_, 1, \_)$$

$$B4 \longrightarrow 2.1$$

$$B5 \longrightarrow 3.4$$

UPV - EHU

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$$
\begin{array}{lcl}
\text{B1} & \longrightarrow & 5.6 \\
\text{B2} & \longrightarrow & 4.9 \\
\text{B3} & \longrightarrow & 9.3 \\
\text{B4} & \longrightarrow & 2.1 \\
\text{B5} & \longrightarrow & 3.4
\end{array}
$$

$$(\_, \_, \_, 1, 2)$$

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$$B1 \longrightarrow 5.6$$

$$B2 \longrightarrow 4.9$$

$$B3 \longrightarrow 9.3 \qquad (\_, \_, \_, 1, 2)$$

$$B4 \longrightarrow 2.1$$

$$B5 \longrightarrow 3.4$$

UPV - EHU

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$$\begin{array}{lll} \text{B1} & \longrightarrow & 5.6 \\ \text{B2} & \longrightarrow & 4.9 \\ \text{B3} & \longrightarrow & 9.3 \\ \text{B4} & \longrightarrow & 2.1 \\ \text{B5} & \longrightarrow & 3.4 \end{array}$$

$$(4, 3, 5, 1, 2)$$

Thurstone Order Statistic Models

Ranking biscuits in relation with its sweetness

$$\begin{array}{llll} B1 & \longrightarrow & 5.6 & \sim X_1 \\ B2 & \longrightarrow & 4.9 & \sim X_2 \\ B3 & \longrightarrow & 9.3 & \sim X_3 \\ B4 & \longrightarrow & 2.1 & \sim X_4 \\ B5 & \longrightarrow & 3.4 & \sim X_5 \end{array} \qquad (4, 3, 5, 1, 2)$$

Thurstone Order Statistic Models

Basics

- Each item is associated with a true continue value: sweetness of a cookie, loudness of a sound, etc.
- A judge assesses the cookies or the sounds and classifies them
- Errors are produced because of the lack of exactness of the sensorial apparatus of the judge
- The output of the assessment is a classification of the items

UPV - EHU

Thurstone Order Statistic Models

Basics

- Codify a permutation as a real-valued vector (random keys)
- Given a real vector $(x_1, x_2, \ldots, x_n)$ of length n, a permutation can be obtained by ranking the positions using the values x_i $(i = 1, \ldots, n)$
- Given

$$(2.35, 3.42, 9.35, 0.32, 11.54, 10.42, 5.23, 4.2, 7.8)$$

the permutation obtained is (2 3 7 1 9 8 5 4 6)

UPV - EHU

Thurstone Order Statistic Models

Definition

Given $\{X_1, X_2, \ldots, X_n\}$ random variables with a continuous joint distribution $F(x_1, \ldots, x_n)$, we can define a random ranking σ in such a way that $\sigma(i)$ is the rank that X_i occupied between $X_1, X_2, \ldots, X_n$. In this way:

$$P(\sigma) = P(X_{\sigma^{-1}(1)} < X_{\sigma^{-1}(2)} < \ldots < X_{\sigma^{-1}(n)})$$

Unconstrained X_i's (and therefore F) produce all kind of distributions over S_n

UPV - EHU

Thurstone Models

n-1 parameter model

- Assumption: All the X_i are independent (it produces a proper subset of the distributions over permutations)
- All the distributions are similar $f_i(x) = f(x - \mu_i)$
- The parameters of the model are $(\mu_2 - \mu_1, \ldots, \mu_n - \mu_1)$
- Most common models: f Gaussian or f Gumbel (Luce's Model)

UPV - EHU

Thurstone Models

Learning I

- MLE of the parameters: complex numerical integral problem:

$$\begin{aligned} p(\sigma) &= P(X_{\sigma^{-1}(1)} < X_{\sigma^{-1}(2)} < \ldots < X_{\sigma^{-1}(n)}) \\ &= \int_{\Omega} f(x_1, x_2, \ldots, x_n) dx_1 \cdots dx_n \end{aligned}$$

where $\Omega = \{(x_1, x_2, \ldots, x_n) | x_{\sigma^{-1}(1)} < x_{\sigma^{-1}(2)} < \ldots < x_{\sigma^{-1}(n)}$ with $x_i \in \mathcal{R} \;\; i = 1, \ldots, n\}$

Thurstone Models

Learning Approaches

- Böckenholt (1993): Numerical integration ($n \leq 4$)
- Yao & Böckenholt (1999): Bayesian Gibbs sampling ($n \leq 10$)
- Limited information (Maydeu-Olivares, 1999; 2001; 2003): pair, triplets and tetrads comparisons frequencies ($n \leq 10$)

Thurstone Models

Sampling

- It depends on the complexity of $F(x_1, \ldots, x_n)$

TrueSkill^{TM} (Herbrich et al. 2007)

Outline of the talk

UPV - EHU

Plackett-Luce Model

Ranking objects with Plackett-Luce

O1

O2

O3

O4

O5

Plackett-Luce Model

Ranking objects with Plackett-Luce

O1

O2

O3

O4

O5

Plackett-Luce Model

Ranking objects with Plackett-Luce

O1 $\longrightarrow$ w_1

O2

O3

O4

O5

Plackett-Luce Model

Ranking objects with Plackett-Luce

O1 $\longrightarrow$ w_1

O2 $\longrightarrow$ w_2

O3 $\longrightarrow$ w_3

O4 $\longrightarrow$ w_4

O5 $\longrightarrow$ w_5

UPV - EHU

Plackett-Luce Model

Ranking objects with Plackett-Luce

O1 $\longrightarrow$ w_1 $\frac{w_1}{w_1+w_2+w_3+w_4+w_5}$

O2 $\longrightarrow$ w_2

O3 $\longrightarrow$ w_3

O4 $\longrightarrow$ w_4

O5 $\longrightarrow$ w_5

Plackett-Luce Model

Ranking objects with Plackett-Luce

$$O1 \longrightarrow w_1 \qquad \frac{w_1}{w_1+w_2+w_3+w_4+w_5}$$

$$O2 \longrightarrow w_2 \qquad \frac{w_2}{w_1+w_2+w_3+w_4+w_5}$$

$$O3 \longrightarrow w_3 \qquad \frac{w_3}{w_1+w_2+w_3+w_4+w_5}$$

$$O4 \longrightarrow w_4 \qquad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}$$

$$O5 \longrightarrow w_5 \qquad \frac{w_5}{w_1+w_2+w_3+w_4+w_5}$$

Plackett-Luce Model

Ranking objects with Plackett-Luce

O1	$\longrightarrow$	w_1	$\frac{w_1}{w_1+w_2+w_3+w_4+w_5}$
O2	$\longrightarrow$	w_2	$\frac{w_2}{w_1+w_2+w_3+w_4+w_5}$
O3	$\longrightarrow$	w_3	$\frac{w_3}{w_1+w_2+w_3+w_4+w_5}$
O4	$\longrightarrow$	w_4	$\frac{w_4}{w_1+w_2+w_3+w_4+w_5}$
O5	$\longrightarrow$	w_5	$\frac{w_5}{w_1+w_2+w_3+w_4+w_5}$

$$(\_,\_,\_,\_,\_)$$

UPV - EHU

Plackett-Luce Model

Ranking objects with Plackett-Luce

$$O1 \longrightarrow w_1 \qquad \frac{w_1}{w_1+w_2+w_3+w_4+w_5}$$

$$O2 \longrightarrow w_2 \qquad \frac{w_2}{w_1+w_2+w_3+w_4+w_5}$$

$$O3 \longrightarrow w_3 \qquad \frac{w_3}{w_1+w_2+w_3+w_4+w_5} \qquad (\_,\_,\_,\_,\_)$$

$$\color{red}{O4 \longrightarrow w_4 \qquad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}}$$

$$O5 \longrightarrow w_5 \qquad \frac{w_5}{w_1+w_2+w_3+w_4+w_5}$$

UPV - EHU

Plackett-Luce Model

Ranking objects with Plackett-Luce

$$O1 \longrightarrow w_1 \quad \frac{w_1}{w_1+w_2+w_3+w_4+w_5}$$

$$O2 \longrightarrow w_2 \quad \frac{w_2}{w_1+w_2+w_3+w_4+w_5}$$

$$O3 \longrightarrow w_3 \quad \frac{w_3}{w_1+w_2+w_3+w_4+w_5}$$

$$O4 \longrightarrow w_4 \quad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}$$

$$O5 \longrightarrow w_5 \quad \frac{w_5}{w_1+w_2+w_3+w_4+w_5}$$

$$(\_,\_,\_,1,\_)$$

Plackett-Luce Model

Ranking objects with Plackett-Luce

$$O1 \longrightarrow w_1 \quad \frac{w_1}{w_1+w_2+w_3+w_4+w_5}$$

$$O2 \longrightarrow w_2 \quad \frac{w_2}{w_1+w_2+w_3+w_4+w_5}$$

$$O3 \longrightarrow w_3 \quad \frac{w_3}{w_1+w_2+w_3+w_4+w_5}$$

$$\cancel{O4} \longrightarrow w_4 \quad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}$$

$$O5 \longrightarrow w_5 \quad \frac{w_5}{w_1+w_2+w_3+w_4+w_5}$$

$$(\_, \_, \_, 1, \_)$$

UPV - EHU

Plackett-Luce Model

Ranking objects with Plackett-Luce

$$O1 \longrightarrow w_1 \quad \frac{w_1}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$O2 \longrightarrow w_2 \quad \frac{w_2}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$O3 \longrightarrow w_3 \quad \frac{w_3}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$\cancel{O4} \longrightarrow w_4 \quad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}$$

$$O5 \longrightarrow w_5 \quad \frac{w_5}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$(\_, \_, \_, 1, \_)$$

UPV - EHU

Plackett-Luce Model

Ranking objects with Plackett-Luce

$$O1 \longrightarrow w_1 \quad \frac{w_1}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$O2 \longrightarrow w_2 \quad \frac{w_2}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$O3 \longrightarrow w_3 \quad \frac{w_3}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$\cancel{O4} \longrightarrow w_4 \quad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}$$

$$O5 \longrightarrow w_5 \quad \frac{w_5}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$(\_, \_, \_, 1, \_)$$

Plackett-Luce Model

Ranking objects with Plackett-Luce

$$O1 \longrightarrow w_1 \quad \frac{w_1}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$O2 \longrightarrow w_2 \quad \frac{w_2}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$O3 \longrightarrow w_3 \quad \frac{w_3}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$\cancel{O4} \longrightarrow w_4 \quad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}$$

$$O5 \longrightarrow w_5 \quad \frac{w_5}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$(\_, \_, \_, 1, 2)$$

Plackett-Luce Model

Ranking objects with Plackett-Luce

$$O1 \longrightarrow w_1 \quad \frac{w_1}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$O2 \longrightarrow w_2 \quad \frac{w_2}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$O3 \longrightarrow w_3 \quad \frac{w_3}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$\cancel{O4} \longrightarrow w_4 \quad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}$$

$$\cancel{O5} \longrightarrow w_5 \quad \frac{w_5}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$(\_,\_,\_,1,2)$$

UPV - EHU

Plackett-Luce Model

Ranking objects with Plackett-Luce

$$O1 \longrightarrow w_1 \qquad \frac{w_1}{w_1+w_2+w_3+\cancel{w_4}+\cancel{w_5}}$$

$$O2 \longrightarrow w_2 \qquad \frac{w_2}{w_1+w_2+w_3+\cancel{w_4}+\cancel{w_5}}$$

$$O3 \longrightarrow w_3 \qquad \frac{w_3}{w_1+w_2+w_3+\cancel{w_4}+\cancel{w_5}} \qquad (\_,\_,\_,1,2)$$

$$\cancel{O4} \longrightarrow w_4 \qquad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}$$

$$\cancel{O5} \longrightarrow w_5 \qquad \frac{w_5}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

UPV - EHU

Plackett-Luce Model

Ranking objects with Plackett-Luce

$O1 \longrightarrow w_1 \qquad \frac{w_1}{w_1+w_2+w_3+\cancel{w_4}+\cancel{w_5}}$

$O2 \longrightarrow w_2 \qquad \frac{w_2}{w_1+w_2+w_3+\cancel{w_4}+\cancel{w_5}}$

$O3 \longrightarrow w_3 \qquad \frac{w_3}{w_1+w_2+w_3+\cancel{w_4}+\cancel{w_5}}$

$\cancel{O4} \longrightarrow w_4 \qquad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}$

$\cancel{O5} \longrightarrow w_5 \qquad \frac{w_5}{w_1+w_2+w_3+\cancel{w_4}+w_5}$

$(4, 3, 5, 1, 2)$

UPV - EHU

Plackett-Luce Model

Ranking objects with Plackett-Luce

$$O1 \longrightarrow w_1 \qquad \frac{w_1}{w_1+w_2+w_3+\cancel{w_4}+\cancel{w_5}}$$

$$O2 \longrightarrow w_2 \qquad \frac{w_2}{w_1+w_2+w_3+\cancel{w_4}+\cancel{w_5}}$$

$$O3 \longrightarrow w_3 \qquad \frac{w_3}{w_1+w_2+w_3+\cancel{w_4}+\cancel{w_5}} \qquad (4,3,5,1,2)$$

$$\cancel{O4} \longrightarrow w_4 \qquad \frac{w_4}{w_1+w_2+w_3+w_4+w_5}$$

$$\cancel{O5} \longrightarrow w_5 \qquad \frac{w_5}{w_1+w_2+w_3+\cancel{w_4}+w_5}$$

$$p(4,3,5,1,2) = \frac{w_4}{w_1+w_2+w_3+w_4+w_5} \cdot \frac{w_5}{w_1+w_2+w_3+w_5} \cdot \frac{w_2}{w_1+w_2+w_3} \cdot \frac{w_1}{w_1+w_3}$$

Plackett-Luce Model

Definition procedure

- Parameters: $\mathbf{W} = (w_1, w_2, \ldots, w_n)$ with $\sum_{i=1}^{n} w_i = 1$ and $w_i > 0$
- w_i is the probability of chosen object i
- Procedure:
  1. An object is chosen using **W**
  2. Update **W** and chose another object

Model

$$P(\sigma) = \prod_{i=1}^{n-1} \frac{w_{\sigma^{-1}(i)}}{\sum_{j=i}^{n} w_{\sigma^{-1}(j)}}$$

UPV - EHU

Plackett-Luce Model

Properties

- The choice probability ratio between two items is independent of any other items in the set
- Easy to extend to partial rankings:

$$P(\sigma) = \prod_{i=1}^{n'-1} \frac{w_{\sigma^{-1}(i)}}{\sum_{j=i}^{n'} w_{\sigma^{-1}(j)}}$$

- Plackett-Luce model can be seen as a Thurstone model with f Gumbel

UPV - EHU

Plackett-Luce model

Learning

- Given a dataset $\{\sigma^1, \sigma^2, \ldots, \sigma^N\}$ find **W**
- Two main approaches have been developed
  - MM algorithm (Hunter, 2004)
  - Bayesian approach by means of message passing (Guiver and Snelson, 2009)

UPV - EHU

P-L Learning

Minorization-Maximization (MM) algorithm

- It is an iterative algorithm to calculate ML parameters
- EM algorithm can be considered as a special case of MM
- The key idea: Find a function $Q(\mathbf{w}|\mathbf{w}`)$ such that (minorization):

$$Q(\mathbf{w}|\mathbf{w}`) \leq L(\mathbf{w})$$

  with equality if $\mathbf{w} = \mathbf{w}`$
- In this case it happens that:

$$\text{if} \quad Q(\mathbf{w}|\mathbf{w}') \geq Q(\mathbf{w}'|\mathbf{w}') \;\; \text{then} \quad L(\mathbf{w}) \geq L(\mathbf{w}')$$

UPV - EHU

P-L Learning

Minorization-Maximization (MM) algorithm

The algorithm is as follows:

1. Give an initial guess for the parameters $\mathbf{w}^0$
2. Set $l = 0$
3. Construct function $Q(\mathbf{w}|\mathbf{w}^l)$
4. Find $\mathbf{w}^{l+1} = argmax_{\mathbf{w}} Q(\mathbf{w}|\mathbf{w}^l)$
5. If $||\mathbf{w}^{l+1} - \mathbf{w}^l|| < \theta$ stop. Otherwise set $l = l + 1$ and go to step 3

UPV - EHU

P-L Learning

Minorization-Maximization (MM) algorithm

The algorithm is as follows:

1. Give an initial guess for the parameters $\mathbf{w}^0$
2. Set $l = 0$
3. Construct function $Q(\mathbf{w}|\mathbf{w}^l)$
4. Find $\mathbf{w}^{l+1} = argmax_{\mathbf{w}} Q(\mathbf{w}|\mathbf{w}^l)$
5. If $||\mathbf{w}^{l+1} - \mathbf{w}^l|| < \theta$ stop. Otherwise set $l = l + 1$ and go to step 3

How to construct $Q(\mathbf{w}|\mathbf{w}^l)$

UPV - EHU

P-L Learning

Minorization-Maximization (MM) algorithm

The algorithm is as follows:

1. Give an initial guess for the parameters $\mathbf{w}^0$
2. Set $l = 0$
3. Construct function $Q(\mathbf{w}|\mathbf{w}^l)$
4. Find $\mathbf{w}^{l+1} = argmax_{\mathbf{w}} Q(\mathbf{w}|\mathbf{w}^l)$
5. If $||\mathbf{w}^{l+1} - \mathbf{w}^l|| < \theta$ stop. Otherwise set $l = l + 1$ and go to step 3

How to construct $Q(\mathbf{w}|\mathbf{w}^l)$

How to calculate $argmax_{\mathbf{w}} Q(\mathbf{w}|\mathbf{w}^l)$

P-L Learning

MM for P-L

Maximum Likelihood function given a sample $\{\sigma^1, \ldots, \sigma^N\}$:

$$\ln L(\mathbf{w}|\sigma^1, \ldots, \sigma^N) = \sum_{k=1}^{N} \sum_{i=1}^{n-1} \left(\ln w_{(\sigma^k)^{-1}(i)} - \ln \sum_{j=i}^{n} w_{(\sigma^k)^{-1}(j)} \right)$$

Constructing Q

$$\forall x, y > 0 \quad \text{we have} \quad -\ln x \geq 1 - \ln y - \frac{x}{y}$$

Taking y in the previous formula as $\mathbf{w}^l$ and applying it to L we obtain Q

P-L Learning

MM for P-L. Function Q

$$Q(\mathbf{w}|\mathbf{w}^l) = \sum_{k=1}^{N} \sum_{i=1}^{n-1} \left(\ln w_{(\sigma^k)^{-1}(i)} - \frac{\sum_{j=i}^{n} w_{(\sigma^k)^{-1}(j)}}{\sum_{j=i}^{n} w^l_{(\sigma^k)^{-1}(j)}} \right)$$

P-L Learning

MM for P-L. Maximize Q

We maximize Q by deriving and equating to 0:

$$w_r^{l+1} = \frac{h_r}{\sum_{k=1}^{N} \sum_{i=1}^{n-1} \delta_{kir} [\sum_{j=i}^{n} w^l_{(\sigma^k)^{-1}(j)}]^{-1}}$$

where h_r is the number of rankings in which the $r-$th item is ranked higher than last and

$$\delta_{kir} = \begin{cases} 1 & \text{if an item higher than } i-1 \text{ is ranked } r \\ & \text{in the } k\text{-th permutation} \\ 0 & \text{otherwise} \end{cases}$$

P-L Learning

Comments on the MM approach

- To guarantee convergence several properties have to be complied with. For instance:

  > *...in every possible partition of the items into two nonempty subsets, some item in the second set ranks higher than some item in the first set at least once..*

- May be applied to partial rankings without any changes
- May be combined with Newton-Raphson method
- Hunter (2004) reports experiments with up to $n = 80$

P-L Learning

Bayesian Approach (Guiver and Snelson, 2009)

- Avoid the overfitting that ML parameters can suffer in certain situations
- Give a method that may be globally applied without the constraints of the MM method
- Bayesian approach: given the dataset, assume a priori distribution over the parameters and carry out the inference that will obtain the a posteriori distribution

P-L Learning

Bayesian Approach (Guiver and Snelson, 2009)

- The authors assume a Gamma distribution as a prior:

$$w_i \sim Gamma(v|\alpha_0, \beta_0)$$

- Assume a full factorization for the posterior $p(\mathbf{w}) \approx \prod_{i=1}^{n} p(w_i)$
- Use a message-passing algorithm (power Expectation-Propagation) to carry out inference

P-L Inference

Inference

- Sampling is trivial in this model
- Marginal calculation can be exponential: $P(\sigma^{-1}(n) = i)$

Machine Learning Applications

ML Applications

- Cheng et al. (2010): Label ranking
- Chen et al. (2007): Document ranking

Outline of the talk

UPV - EHU

Ranking induced by pair-comparisons

Definition

- Babington-Smith proposes a model based on pair comparisons:

$$p_{i,j} = \text{probability of preferring item } i \text{ to item } j$$

  if only that comparison were to be made
- Assuming no ties, for n objects there are $\binom{n}{2}$ possible comparisons

UPV - EHU

Ranking induced by pair-comparisons

Definition

- Given an ordering it is easy to get the pair comparisons:

$$(312) \Rightarrow 3 > 1, \;\; 3 > 2, \;\; 1 > 2$$

- The opposite is not true in general:

$$3 > 1, \;\; 1 > 2, \;\; 2 > 3$$

Ranking induced by pair-comparisons

Definition

- A ranking is obtained carrying out all kind of comparisons until a consistent set of comparisons is obtained:

$$P(\sigma) \propto \prod_{(i,j)|\sigma^{-1}(i)<\sigma^{-1}(j)} p_{i,j}$$

- For instance given the *ordering* (1 3 4 2):

$$p(1\ 3\ 4\ 2) \propto p_{1,3} \cdot p_{1,4} \cdot p_{1,2} \cdot p_{3,4} \cdot p_{3,2} \cdot p_{4,2}$$

- The number of parameters is quadratic $n(n-1)/2$
- There is no closed form for the normalization constant
- Simplifications....

Ranking induced by pair-comparisons

Mallows-Bradley-Terry models

- The model comes defined by a vector of weights $(v_1, \ldots, v_n)$ each one associated with an item:

$$p_{i,j} = \frac{v_i}{v_i + v_j} \text{ with } \sum_{i=1}^{n} v_i = 1 \text{ and } v_i > 0$$

- MM algorithms have been given to learn the MLE parameters (Hunter, 2004)

UPV - EHU

Ranking induced by pair-comparisons

Extensions to Mallows-Bradley-Terry models

- Agresti (1990) considers a model where the probability of i beating j depends on which individual is listed first:

$$p_{ij} = \begin{cases} \theta w_i/(\theta v_i + v_j) & \text{if } i \text{ is home} \\ v_i/(v_i + \theta v_j) & \text{if } j \text{ is home} \end{cases}$$

- Rao and Kupper (1967) allow ties:

$$\begin{aligned} P(i \text{ beats } j) &= v_i/(v_i + \theta v_j) \\ P(j \text{ beats } i) &= v_j/(\theta v_i + v_j) \\ P(j \text{ ties } i) &= (\theta^2 - 1) v_i v_j/((\theta v_i + v_j)(v_i + \theta v_j)) \end{aligned}$$

Outline of the talk

Distance-based ranking models (Mallows models)

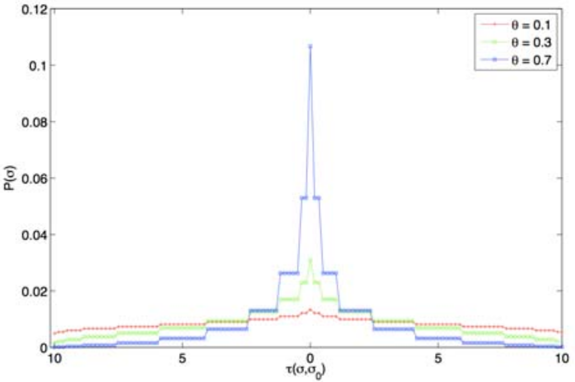
Definition

$$p(\sigma|\theta,\sigma_0) = \frac{1}{Z(\theta)} e^{-\theta d(\sigma,\sigma_0)}$$

- It is an exponential model
- d is a distance between permutations such that:
  - $d(\sigma,\pi) \geq 0 \;\; \forall \sigma, \pi$ with equality iff $\sigma = \pi$
  - Right-invariant property: $d(\sigma,\pi) = d(\sigma\phi,\pi\phi) \;\; \forall \sigma,\pi,\phi$
- Two parameters:
  - Central permutation σ_0
  - Spread parameter $\theta \geq 0$
- $Z(\theta)$ is the partition function

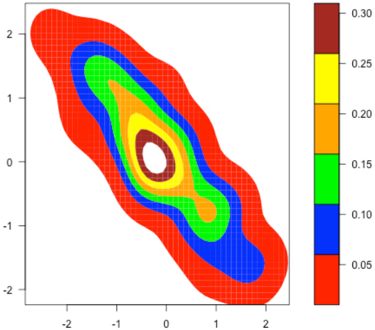
Mallows model

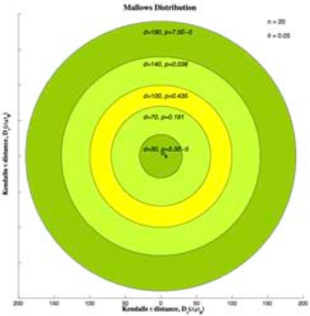
Equivalent to Gaussian distribution for permutations



Mallows model

Equivalent to Gaussian distribution for permutations





UPV - EHU

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi, \sigma) = \sum_{i<j} I\{(\pi(i) - \pi(j))(\sigma(i) - \sigma(j)) < 0\}$$

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi,\sigma) = \sum_{i<j} I\{(\pi(i) - \pi(j))(\sigma(i) - \sigma(j)) < 0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$		$\sigma = (1\ 3\ 4\ 2\ 5)$		T
$(1,2)$	$\pi(1) = 2$	$\pi(2) = 3$	$\sigma(1) = 1$	$\sigma(2) = 3$	

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi,\sigma)=\sum_{i<j} I\{(\pi(i)-\pi(j))(\sigma(i)-\sigma(j))<0\}$$

Example

	$\pi=(2\ 3\ 1\ 5\ 4)$	$\sigma=(1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1)=2<\pi(2)=3$	$\sigma(1)=1 \quad \sigma(2)=3$	

UPV - EHU

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi,\sigma) = \sum_{i<j} I\{(\pi(i)-\pi(j))(\sigma(i)-\sigma(j)) < 0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$	$\sigma = (1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1) = 2 < \pi(2) = 3$	$\sigma(1) = 1 < \sigma(2) = 3$	

UPV - EHU

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi,\sigma)=\sum_{i<j} I\{(\pi(i)-\pi(j))(\sigma(i)-\sigma(j))<0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$	$\sigma = (1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1)=2<\pi(2)=3$	$\sigma(1)=1<\sigma(2)=3$	0

UPV - EHU

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi,\sigma) = \sum_{i<j} I\{(\pi(i)-\pi(j))(\sigma(i)-\sigma(j))<0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$	$\sigma = (1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1)=2 < \pi(2)=3$	$\sigma(1)=1 < \sigma(2)=3$	0
$(1,3)$	$\pi(1)=2 \quad \pi(3)=1$	$\sigma(1)=1 \quad \sigma(3)=4$	

UPV - EHU

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi, \sigma) = \sum_{i<j} I\{(\pi(i) - \pi(j))(\sigma(i) - \sigma(j)) < 0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$	$\sigma = (1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1) = 2 < \pi(2) = 3$	$\sigma(1) = 1 < \sigma(2) = 3$	0
$(1,3)$	$\pi(1) = 2 > \pi(3) = 1$	$\sigma(1) = 1 \quad \sigma(3) = 4$	

UPV - EHU

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi,\sigma) = \sum_{i<j} I\{(\pi(i)-\pi(j))(\sigma(i)-\sigma(j)) < 0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$	$\sigma = (1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1) = 2 < \pi(2) = 3$	$\sigma(1) = 1 < \sigma(2) = 3$	0
$(1,3)$	$\pi(1) = 2 > \pi(3) = 1$	$\sigma(1) = 1 < \sigma(3) = 4$	

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi,\sigma) = \sum_{i<j} I\{(\pi(i)-\pi(j))(\sigma(i)-\sigma(j)) < 0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$	$\sigma = (1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1) = 2 < \pi(2) = 3$	$\sigma(1) = 1 < \sigma(2) = 3$	0
$(1,3)$	$\pi(1) = 2 > \pi(3) = 1$	$\sigma(1) = 1 < \sigma(3) = 4$	1

UPV - EHU

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi, \sigma) = \sum_{i<j} I\{(\pi(i) - \pi(j))(\sigma(i) - \sigma(j)) < 0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$	$\sigma = (1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1) = 2 < \pi(2) = 3$	$\sigma(1) = 1 < \sigma(2) = 3$	0
$(1,3)$	$\pi(1) = 2 > \pi(3) = 1$	$\sigma(1) = 1 < \sigma(3) = 4$	1
$\vdots$	$\vdots$ $\vdots$	$\vdots$	$\vdots$

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi,\sigma) = \sum_{i<j} I\{(\pi(i)-\pi(j))(\sigma(i)-\sigma(j)) < 0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$	$\sigma = (1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1) = 2 < \pi(2) = 3$	$\sigma(1) = 1 < \sigma(2) = 3$	0
$(1,3)$	$\pi(1) = 2 > \pi(3) = 1$	$\sigma(1) = 1 < \sigma(3) = 4$	1
$\vdots$	$\vdots$ $\vdots$	$\vdots$	$\vdots$
$(4,5)$	$\pi(4) = 5 \quad \pi(5) = 4$	$\sigma(4) = 2 \quad \sigma(5) = 5$	

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi, \sigma) = \sum_{i<j} I\{(\pi(i) - \pi(j))(\sigma(i) - \sigma(j)) < 0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$	$\sigma = (1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1) = 2 < \pi(2) = 3$	$\sigma(1) = 1 < \sigma(2) = 3$	0
$(1,3)$	$\pi(1) = 2 > \pi(3) = 1$	$\sigma(1) = 1 < \sigma(3) = 4$	1
$\vdots$	$\vdots$ $\vdots$	$\vdots$	$\vdots$
$(4,5)$	$\pi(4) = 5 > \pi(5) = 4$	$\sigma(4) = 2$ $\sigma(5) = 5$	

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi,\sigma)=\sum_{i<j} I\{(\pi(i)-\pi(j))(\sigma(i)-\sigma(j))<0\}$$

Example

	$\pi=(2\ 3\ 1\ 5\ 4)$	$\sigma=(1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1)=2<\pi(2)=3$	$\sigma(1)=1<\sigma(2)=3$	0
$(1,3)$	$\pi(1)=2>\pi(3)=1$	$\sigma(1)=1<\sigma(3)=4$	1
$\vdots$	$\vdots$ $\vdots$	$\vdots$	$\vdots$
$(4,5)$	$\pi(4)=5>\pi(5)=4$	$\sigma(4)=2<\sigma(5)=5$	

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi, \sigma) = \sum_{i<j} I\{(\pi(i) - \pi(j))(\sigma(i) - \sigma(j)) < 0\}$$

Example

	$\pi = (2\ 3\ 1\ 5\ 4)$	$\sigma = (1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1) = 2 < \pi(2) = 3$	$\sigma(1) = 1 < \sigma(2) = 3$	0
$(1,3)$	$\pi(1) = 2 > \pi(3) = 1$	$\sigma(1) = 1 < \sigma(3) = 4$	1
$\vdots$	$\vdots$ $\vdots$	$\vdots$	$\vdots$
$(4,5)$	$\pi(4) = 5 > \pi(5) = 4$	$\sigma(4) = 2 < \sigma(5) = 5$	1

Distances between permutations I

- Kendall-tau metric. Measures the number of disagreements between two permutations:

$$T(\pi,\sigma)=\sum_{i<j} I\{(\pi(i)-\pi(j))(\sigma(i)-\sigma(j))<0\}$$

Example

	$\pi=(2\ 3\ 1\ 5\ 4)$	$\sigma=(1\ 3\ 4\ 2\ 5)$	T
$(1,2)$	$\pi(1)=2<\pi(2)=3$	$\sigma(1)=1<\sigma(2)=3$	0
$(1,3)$	$\pi(1)=2>\pi(3)=1$	$\sigma(1)=1<\sigma(3)=4$	1
$\vdots$	$\vdots$ $\vdots$	$\vdots$	$\vdots$
$(4,5)$	$\pi(4)=5>\pi(5)=4$	$\sigma(4)=2<\sigma(5)=5$	1

$$T((2\ 3\ 1\ 5\ 4),(1\ 3\ 4\ 2\ 5))=5$$

- EHU

Distances between permutations II

- Kendall-tau is equivalente to: minimum number of adjacent swaps to go from π^{-1} to σ^{-1}.
- $0 \leq T(\pi, \sigma) \leq n(n-1)/2$

Distances between permutations III

- Spearman rho metric:

$$R(\pi,\sigma) = \left(\sum_{i=1}^{n}(\pi(i)-\sigma(i))^2\right)^{1/2}$$

- Spearman footrule:

$$F(\pi,\sigma) = \sum_{i=1}^{n}|\pi(i)-\sigma(i)|$$

Distances between permutations IV

- Hamming distance:

$$H(\pi, \sigma) = \sum_{i=1}^{n} I\{\pi(i) \neq \sigma(i)\}$$

- Cayley metric:

$$C(\pi, \sigma) = \text{minimum number of transpositions to go from } \pi \text{ to } \sigma$$

Distances between permutations V

- Ulam metric

$$U(\pi,\sigma) = n - \text{maximum number of items ranked the same relative order by } \pi \text{ and } \sigma$$

- Example:

$$e = (1\ 2\ 3\ 4\ 5) \qquad \pi = (1\ 4\ 3\ 2\ 5)$$

Distances between permutations V

- Ulam metric

$$U(\pi, \sigma) = n - \text{maximum number of items ranked the same relative order by } \pi \text{ and } \sigma$$

- Example:

$$e = (1\ 2\ 3\ 4\ 5) \qquad \pi = (1\ 4\ 3\ 2\ 5)$$

$$e = (1\ 2\ 3\ 4\ 5) \qquad \pi = ({\color{red}1}\ 4\ {\color{red}3}\ 2\ {\color{red}5})$$

UPV - EHU

Distances between permutations V

- Ulam metric

$$U(\pi, \sigma) = n - \text{maximum number of items ranked the same relative order by } \pi \text{ and } \sigma$$

- Example:

$$e = (1\ 2\ 3\ 4\ 5) \qquad \pi = (1\ 4\ 3\ 2\ 5)$$

$$e = (1\ 2\ 3\ 4\ 5) \qquad \pi = ({\color{red}1}\ 4\ {\color{red}3}\ 2\ {\color{red}5})$$

$$U(e, \pi) = 5 - 3 = 2$$

UPV - EHU

Distances between permutations V

- Ulam metric

$$U(\pi, \sigma) = n - \text{maximum number of items ranked the same relative order by } \pi \text{ and } \sigma$$

- Example:

$$e = (1\ 2\ 3\ 4\ 5) \qquad \pi = (1\ 4\ 3\ 2\ 5)$$

$$e = (1\ 2\ 3\ 4\ 5) \qquad \pi = ({\color{red}1}\ 4\ {\color{red}3}\ 2\ {\color{red}5})$$

$$U(e, \pi) = 5 - 3 = 2$$

- It is equivalent to the minimum number of "delete-shift-insert" operations to go from π to σ

Mallows model

Learning

- Given a dataset $\{\sigma^1, \sigma^2, \ldots, \sigma^N\}$ find σ_0 and θ
- Maximum likelihood estimation
  - $\hat{\sigma}_0 = arg\ min_\sigma \sum_{i=1}^N d(\sigma^i, \sigma)$ (consensus ranking, NP-hard for T)
  - $\hat{\theta}$ can be found by a standard numerical method for convex optimization, given $\hat{\sigma}_0$

Inference

- For all distance d, Gibbs sampling
- There are alternative methods depending on d

Mallows model with Kendall distance

Definition

- In this particular case the partition function can be calculated in a closed form:

$$p(\sigma|\theta,\sigma_0) = \frac{1}{Z(\theta)} e^{-\theta d(\sigma,\sigma_0)}$$

$$p(\sigma|\theta,\sigma_0) = \frac{(1-e^{-\theta})^{n-1}}{\prod_{i=1}^{n-1}(1-e^{-(n-i+1)\theta})} e^{-\theta T(\sigma,\sigma_0)}$$

- Learning. Kemeny ranking problem: $\hat{\sigma}_0 = arg\ min_\sigma \sum_{i=1}^{N} T(\sigma_i,\sigma)$

Mallows model with Kendall distance

Learning

- Ali and Meila (2012) compare 104 algorithms in the Kemeny ranking problem. *Borda* one of the best:
  1. Calculate for all index i:

$$v_i = 1/N \sum_{j=1}^{N} \sigma^j(i)$$

  2. Assign $\hat{\sigma}_0^{-1}(1)$ with the index of the smallest $\{v_1, \ldots, v_n\}$, $\hat{\sigma}_0^{-1}(2)$ with second smallest and so on

Mallows model

Problems

- Unimodality
- Estimation
- Permutations at the same distance from σ_0 have the same probability

Mallows model

Extensions

- Mixture of Mallows models (Murphy and Martin, 2003; Lee and Yu, 2012; Lu and Boutilier, 2011)
- Two-side infinite extension (Gnedin and Grigori Olshanski, 2012)
- Bao and Meila (2010) extends the Mallows model to infinite items. They learn the model from top-t lists of items and extend it to multi-modal data
- Generalized Mallows models

Generalized Mallows models

Definition

- Fligner and Verducci (1986): if a distance d can be written as:

$$d(\pi, \sigma) = \sum_{i=1}^{n} S_i(\sigma, \pi)$$

and the S_i are independent under the uniform distribution, then the Mallows model is factorizable. Even more, it can be generalized to a n-parameters model:

$$P(\sigma) \propto exp(-\sum_{j=1}^{n} \theta_j S_j(\sigma, \sigma_0))$$

where a different spread parameter θ_i is associated with each position in the permutation

Generalized Mallows models

Kendall distance

- Kendall distance can be decomposed as:

$$T(\pi, \sigma) = \sum_{i=1}^{n-1} V_i(\pi, \sigma)$$

  where $V_i(\pi, \sigma)$ is defined as follows:

$$\sum_{j=i+1}^{n} I\{(\pi(i) - \pi(j))(\sigma(i) - \sigma(j)) < 0\}$$

- Taking $\pi = e$ there exists a bijection between the set of vectors $(v_1, \ldots, v_{n-1})$ and the set of permutations σ

Generalized Mallows models

Kendall distance

- If we assign a different spread parameter θ_i to each position i then

$$T(\sigma,\sigma_0)=\sum_{i=1}^{n-1}\theta_i V_i(\sigma,\sigma_0)$$

and the model can be written as:

$$p(\sigma)=\prod_{i=1}^{n-1}\frac{1}{\psi_i(\theta_i)}e^{-\theta_i V_i(\sigma,\sigma_0)}=\prod_{i=1}^{n-1}\frac{1-e^{-\theta_i}}{1-e^{-(n-i+1)\theta_i}}e^{-\theta_i V_i(\sigma,\sigma_0)}$$

UPV - EHU

Generalized Mallows models with Kendall distance

Sampling

- The distribution of V_i $(i = 1, \ldots, n-1)$ under the model is known:

$$p(V_i(\sigma, \sigma_0) = r) = \frac{e^{-r\theta_i}}{\psi(\theta_i)}$$

Learning

- Mandhani and Meila (2009) give an A* search algorithm with an admissible heuristic

Generalized Mallows model

Cayley distance

- Define the following X_i variables:

$$
\begin{aligned}
cycle_\sigma(i) &= \{\sigma^k(i) | k = 0, 1, \ldots\} \\
&\quad \text{where } \sigma^0(i) = i \text{ and } \sigma^k(i) = \sigma^{k-1}(\sigma(i))
\end{aligned}
$$

$$
X_i(\sigma) = \begin{cases} 0 & \text{if } i = \max\{cycle_\sigma(i)\} \\ 1 & \text{otherwise} \end{cases}
$$

Generalized Mallows model

Cayley distance

- Define the following X_i variables:

$$
\begin{aligned}
cycle_\sigma(i) &= \{\sigma^k(i) | k = 0, 1, \ldots\} \\
&\quad \text{where } \sigma^0(i) = i \text{ and } \sigma^k(i) = \sigma^{k-1}(\sigma(i))
\end{aligned}
$$

$$
X_i(\sigma) = \begin{cases} 0 & \text{if } i = \max\{cycle_\sigma(i)\} \\ 1 & \text{otherwise} \end{cases}
$$

- Example

$$\sigma = (3\ 1\ 2\ 5\ 4)$$

Generalized Mallows model

Cayley distance

- Define the following X_i variables:

$$\begin{aligned} cycle_\sigma(i) &= \{\sigma^k(i)|k = 0, 1, \ldots\} \\ & \quad \text{where } \sigma^0(i) = i \text{ and } \sigma^k(i) = \sigma^{k-1}(\sigma(i)) \end{aligned}$$

$$X_i(\sigma) = \begin{cases} 0 & \text{if } i = \max\{cycle_\sigma(i)\} \\ 1 & \text{otherwise} \end{cases}$$

- Example

$$\sigma = (3\ 1\ 2\ 5\ 4) \quad 1 \rightarrow 2 \rightarrow 3 \rightarrow 1$$

Generalized Mallows model

Cayley distance

- Define the following X_i variables:

$$\begin{aligned} cycle_\sigma(i) &= \{\sigma^k(i) | k = 0, 1, \ldots\} \\ &\quad \text{where } \sigma^0(i) = i \text{ and } \sigma^k(i) = \sigma^{k-1}(\sigma(i)) \end{aligned}$$

$$X_i(\sigma) = \begin{cases} 0 & \text{if } i = \max\{cycle_\sigma(i)\} \\ 1 & \text{otherwise} \end{cases}$$

- Example

$$\sigma = (3\ 1\ 2\ 5\ 4) \quad 1 \to 2 \to 3 \to 1 \qquad 4 \to 5 \to 4$$

Generalized Mallows model

Cayley distance

- Define the following X_i variables:

$$\begin{aligned} cycle_\sigma(i) &= \{\sigma^k(i) | k = 0, 1, \ldots\} \\ &\quad \text{where } \sigma^0(i) = i \text{ and } \sigma^k(i) = \sigma^{k-1}(\sigma(i)) \end{aligned}$$

$$X_i(\sigma) = \begin{cases} 0 & \text{if } i = \max\{cycle_\sigma(i)\} \\ 1 & \text{otherwise} \end{cases}$$

- Example

$$\sigma = (3\ 1\ 2\ 5\ 4) \quad 1 \rightarrow 2 \rightarrow 3 \rightarrow 1 \qquad 4 \rightarrow 5 \rightarrow 4$$

$$\begin{aligned} cycle_\sigma(1) = cycle_\sigma(2) = cycle_\sigma(3) &= \{1, 2, 3\} \\ cycle_\sigma(4) = cycle_\sigma(4) &= \{4, 5\} \end{aligned}$$

Generalized Mallows model

Cayley distance

- Define the following X_i variables:

$$cycle_\sigma(i) = \{\sigma^k(i) | k = 0, 1, \ldots\}$$
$$\text{where } \sigma^0(i) = i \text{ and } \sigma^k(i) = \sigma^{k-1}(\sigma(i))$$

$$X_i(\sigma) = \begin{cases} 0 & \text{if } i = \max\{cycle_\sigma(i)\} \\ 1 & \text{otherwise} \end{cases}$$

- Example

$$\sigma = (3\ 1\ 2\ 5\ 4) \quad 1 \to 2 \to 3 \to 1 \qquad 4 \to 5 \to 4$$
$$cycle_\sigma(1) = cycle_\sigma(2) = cycle_\sigma(3) = \{1, 2, 3\}$$
$$cycle_\sigma(4) = cycle_\sigma(4) = \{4, 5\}$$

$$X_1(\sigma) = X_2(\sigma) = 1 \;\; X_3(\sigma) = 0 \;\; X_4(\sigma) = 1 \;\; X_5(\sigma) = 0$$

Generalized Mallows model

Cayley distance

- Cayley distance can be written as:

$$C(\pi,\sigma) = C(\pi\sigma^{-1}, \sigma\sigma^{-1}) = C(\pi\sigma^{-1}, e) = \sum_{i=1}^{n-1} X_i(\pi\sigma^{-1})$$

- A *similar* development that with Kendall can be given
  - A key difference: there is not a bijection between the set of binary vectors $(X_1, \ldots, X_{n-1})$ and the set of permutations S_n

UPV - EHU

Multi-stage ranking models

Definition

- The ranking is constructed in a process on $n-1$ stages
- At each step a decision is taken
- A modal raking σ_0 is assumed to exist
- The decision only depends on the number of items left (the stage)

UPV - EHU

Multi-stage ranking models

$$\sigma_0 = (2\ 4\ 3\ 5\ 1) \qquad (\_\ \_\ \_\ \_\ \_)$$

$$\sigma_0^{-1} = (5\ 1\ 3\ 2\ 4)$$

Multi-stage ranking models

$$\sigma_0 = (2\ 4\ 3\ 5\ 1) \qquad (-\ \ -\ \ -\ \ -\ \ -)$$
$$\sigma_0^{-1} = (5\ 1\ 3\ 2\ 4)$$

Probability of each choice at stage 1:

$$\{0.4,\ 0.35,\ 0.15,\ 0.1,\ 0.05\}$$

UPV - EHU

Multi-stage ranking models

$$\sigma_0 = (2\ 4\ 3\ 5\ 1) \qquad (-\ \ -\ \ -\ \ -\ \ -)$$
$$\sigma_0^{-1} = (5\ 1\ 3\ 2\ 4)$$

Probability of each choice at stage 1:

$$\{0.4,\ 0.35,\ 0.15,\ 0.1,\ 0.05\}$$

UPV - EHU

Multi-stage ranking models

$$\sigma_0 = (2\ 4\ 3\ 5\ 1) \qquad (1\ -\ -\ -\ -\ )$$

$$\sigma_0^{-1} = (5\ \cancel{1}\ 3\ 2\ 4)$$

Probability of each choice at stage 1:

$$\{0.4,\ 0.35,\ 0.15,\ 0.1,\ 0.05\}$$

Multi-stage ranking models

$$\sigma_0 = (2\ 4\ 3\ 5\ 1) \qquad (1\ -\ -\ -\ -\ )$$
$$\sigma_0^{-1} = (5\ \cancel{1}\ 3\ 2\ 4)$$

Probability of each choice at stage 2:

$$\{0.35,\ 0.3,\ 0.2,\ 0.15\}$$

Multi-stage ranking models

$$\sigma_0 = (2\ 4\ 3\ 5\ 1) \qquad (1\ -\ -\ -\ -\ )$$
$$\sigma_0^{-1} = (5\ \cancel{1}\ 3\ 2\ 4)$$

Probability of each choice at stage 2:

$$\{0.35,\ 0.3,\ 0.2,\ \textcolor{red}{0.15}\}$$

UPV - EHU

Multi-stage ranking models

$$\sigma_0 = (2\ 4\ 3\ 5\ 1) \qquad (1\ -\ -\ \color{red}{2}\ -\ )$$
$$\sigma_0^{-1} = (5\ \cancel{1}\ 3\ 2\ 4)$$

Probability of each choice at stage 2:

$$\{0.35,\ 0.3,\ 0.2,\ \color{red}{0.15}\}$$

Multi-stage ranking models

$$\sigma_0 = (2\ 4\ 3\ 5\ 1) \qquad (1\ 4\ 5\ 2\ 3\ )$$
$$\sigma_0^{-1} = (5\ \cancel{1}\ 3\ 2\ 4)$$

Probability of each choice at stage 2:

$$\{0.35,\ 0.3,\ 0.2,\ 0.15\}$$

UPV - EHU

Multi-stage ranking models

Definition

- It has $n(n-1)/2$ parameters:

  $p(m, r) =$ probability of taken the m-th best decision at stage r

- Let $V_r = m$ if the $(m+1)$-th best decision is taken at stage r. Then

$$P(\pi) = \prod_{r=1}^{n-1} p(V_r, r) \quad \text{factorable model}$$

- Mallows and Plackett-Luce can be seen as multi-stage models

Outline of the talk

UPV - EHU

Models based on marginals

First-order maginals

- First-order marginals $Q = [q_{ij}]$:

$$q_{ij} = P(\sigma(i) = j)$$

- Easy to learn and to represent
- Which distribution is represented by these marginals?

UPV - EHU

Models based on marginals

First-order maginals

- First-order marginals $Q = [q_{ij}]$:

$$q_{ij} = P(\sigma(i) = j)$$

- Easy to learn and to represent
- Which distribution is represented by these marginals?
- Infinite distributions

UPV - EHU

Models based on marginals

First-order marginals

- Criterion to choose between them: The one with the maximum entropy

$$P(\sigma) = exp\left(\sum_{i=1}^{n} Y_{(i,\sigma(i))} - 1\right)$$

where $Y \in \mathbb{R}^{nxn}$

- Obtaining Y is #P-hard (Agrawal et al., 2008)

Models based on marginals

High-order marginals

- Huang et al. (2009) consider models based on high-order marginals
- They use the Fourier transform over the symmetric group to carry out inference tasks
- Irurozki et al. (2011) give a first approach to learning these kind of models

UPV - EHU

Non-parametric models I

Based on Mallows kernels

- Lebanon and Mao (2008) give a probabilistic approach based on Mallows kernels with Kendall distance

$$p(\sigma) = \frac{1}{NZ(\theta)} \sum_{i=1}^{N} e^{-\theta T(\sigma, \sigma^i)}$$

- They were able to learn from partially ranked data:

UPV - EHU

Non-parametric models I

Based on Mallows kernels

- Lebanon and Mao (2008) give a probabilistic approach based on Mallows kernels with Kendall distance

$$p(\sigma) = \frac{1}{NZ(\theta)} \sum_{i=1}^{N} e^{-\theta T(\sigma, \sigma^i)}$$

- They were able to learn from partially ranked data:

$$\sigma = (2\ 3\ 1\ 7\ 6\ 5\ 4\ -\ -\ -\ -\ \ldots)$$

UPV - EHU

Non-parametric models I

Based on Mallows kernels

- Lebanon and Mao (2008) give a probabilistic approach based on Mallows kernels with Kendall distance

$$p(\sigma) = \frac{1}{NZ(\theta)} \sum_{i=1}^{N} e^{-\theta T(\sigma,\sigma^i)}$$

- They were able to learn from partially ranked data:

$$\begin{aligned} \sigma &= (2\;3\;1\;7\;6\;5\;4\;-\;-\;-\;-\;\ldots) \\ C_\sigma &= \{\pi | \pi(i) = \sigma(i)\; i = 1, 2, \ldots, 7\} \end{aligned}$$

Non-parametric models I

Based on Mallows kernels

- Lebanon and Mao (2008) give a probabilistic approach based on Mallows kernels with Kendall distance

$$p(\sigma) = \frac{1}{NZ(\theta)} \sum_{i=1}^{N} e^{-\theta T(\sigma, \sigma^i)}$$

- They were able to learn from partially ranked data:

$$\begin{aligned}
\sigma &= (2\,3\,1\,7\,6\,5\,4\, - - - - \ldots) \\
C_\sigma &= \{\pi | \pi(i) = \sigma(i)\ i = 1, 2, \ldots, 7\} \\
p(\sigma) &= \frac{1}{NZ(\theta)} \sum_{i=1}^{N} \frac{1}{|C_{\sigma^i}|} \sum_{\tau \in C_{\sigma^i}} e^{-\theta T(\sigma, \tau)}
\end{aligned}$$

Non-parametric models I

Based on Mallows kernels

- Lebanon and Mao (2008) give a probabilistic approach based on Mallows kernels with Kendall distance

$$p(\sigma) = \frac{1}{NZ(\theta)} \sum_{i=1}^{N} e^{-\theta T(\sigma,\sigma^i)}$$

- They were able to learn from partially ranked data:

$$\begin{aligned}
\sigma &= (2\ 3\ 1\ 7\ 6\ 5\ 4\ -\ -\ -\ -\ \ldots) \\
C_\sigma &= \{\pi | \pi(i) = \sigma(i)\ i = 1, 2, \ldots, 7\} \\
p(\sigma) &= \frac{1}{NZ(\theta)} \sum_{i=1}^{N} \frac{1}{|C_{\sigma^i}|} \sum_{\tau \in C_{\sigma^i}} e^{-\theta T(\sigma,\tau)}
\end{aligned}$$

Non-parametric models II

Based on Mallows kernels

- Efficient learning
- Lacks: Marginalization, conditioning

Non-parametric models III

Modifying Mallows kernels

- Sun and Lebanon (2012) use a triangular kernel based on Kendall distance to solve problems in recommendation systems
- They can deal with all kind of partial rankings

Based on the Fourier transform

- Barbosa y Kondor (2010) give another kernel approach based on the use of the Fourier transform

Outline of the talk

Full independence - Definition (Huang et al. 2009)

- Independence based on relative rankings
- The subsets $A \subset \{1, \ldots, n\}$ and $B = A^c$ are independent under distribution $p(\sigma)$ if

$$p(\sigma) = p_A(\sigma_A) \cdot p_B(\sigma_B)$$

- Difficult to hold in practice

Full independence - Necessary condition

1 2 5 3 4
2 1 4 3 5
1 2 3 4 5
1 2 3 5 4
1 2 5 4 3
2 1 5 3 4

Each subset of the items maps to a particular subset of the positions

Full independence - Definition (Huang et al. 2009)

- Independence based on relative rankings
- The subsets $A \subset \{1, \ldots, n\}$ and $B = A^c$ are independent under distribution $p(\sigma)$ if

$$p(\sigma) = p_A(\sigma_A) \cdot p_B(\sigma_B)$$

- Difficult to hold in practice

Full independence - Necessary condition

1 2 5 3 4
2 1 4 3 5
1 2 3 4 5
1 2 3 5 4
1 2 5 4 3
2 1 5 3 4

Each subset of the items maps to a particular subset of the positions

Luce model

A ranking σ is built by selecting first selecting the preferred item, then among the remaining items, the second preferred and so on. Given n items, each with weight w_i

$$P(\sigma) = \prod_{k=1}^{n-1} P_{\{i_k,\ldots,i_n\}}(\sigma(k)) = \prod_{k=1}^{n-1} \frac{w_{\sigma^{-1}(k)}}{\sum_{j=k}^{n} w_{\sigma^{-1}(j)}}$$

L-decomposability induces a conditional independence

$P(\sigma^{-1}(k) = i_k | \sigma^{-1}(1) = i_1, \ldots, \sigma^{-1}_{k-1}(k-1) = i_{k-1}) = P_{\{i_k,\ldots,i_n\}}(i_k)$

L-decomposable ranking models

Some Thurstone models and the Mallows model based on Kendall, Hamming or Spearman's footrule are also L-decomposable

Bi-decomposability (Csiszar 2008)

- A ranking model $p(\sigma)$ is bi-decomposable iff $p(\sigma)$ and $p(\sigma^{-1})$ are L-decomposable
- A subset of the L-decomposable models

Bi-decomposable models

- Mallows-Bradley-Terry
- Distance model based on Hamming
- Multi-stage models

TL-decomposability (Csiszar 2009)

- A distribution $p(\sigma)$ is TL-decomposable iff $p(\sigma \circ \pi)$ is L-decomposable for every π
- A subset of the bi-decomposable models
- In other words, for every subset C of the positions the order of these alternatives and the order of the remaining alternatives are independent

TL-decomposition

It can be factored as $p(\sigma) = \prod_{k=1}^{n} c_k(\pi(k))$, $(n-1)^2$ free parameters

TL-decomposable models

- Mallows-Bradley-Terry
- Distance model based on Hamming

First order marginal independence (Huang et al. 2009)

Necessary (not sufficient) condition for full independence
Given this collection of permutations, the frequency matrix is

1 2 5 3 4
2 1 4 3 5
1 2 3 4 5
1 2 3 5 4
1 2 5 4 3
2 1 5 3 4

4	2	0	0	0
2	4	0	0	0
0	0	2	1	3
0	0	3	2	1
0	0	1	3	2

Detection: Finding the independent sets is a bi-clustering problem

Approximation quality: What if every time item 2 is in position 1, item 3 is in position 4?

Context: Tracking problems

First order marginal independence (Huang et al. 2009)

Necessary (not sufficient) condition for full independence
Given this collection of permutations, the frequency matrix is

1 2 5 3 4
2 1 4 3 5
1 2 3 4 5
1 2 3 5 4
1 2 5 4 3
2 1 5 3 4

4	2	0	0	0
2	4	0	0	0
0	0	2	1	3
0	0	3	2	1
0	0	1	3	2

- Detection: Finding the independent sets is a bi-clustering problem
- Approximation quality: What if every time item 2 is in position 1, item 3 is in position 4?
- Context: Tracking problems

Riffle shuffle

relative ordering is maintained

Given two riffle independent subsets

Set A={1, 2, 3} and $p_A(\sigma_A)$
Set B={4, 5, 6} and $p_B(\sigma_B)$

Ranking process of a riffle independent set

$shuffle(\sigma_A, \sigma_B) = shuffle((213), (465)) = [(421635), (421653), \ldots]$

UPV - EHU

Riffle shuffle

relative ordering is maintained

Given two riffle independent subsets

Set A={1, 2, 3} and $p_A(\sigma_A)$
Set B={4, 5, 6} and $p_B(\sigma_B)$

Ranking process of a riffle independent set

$shuffle(\sigma_A, \sigma_B) = shuffle((213), (465)) = [(421635), (421653), \ldots]$

Riffle shuffle

relative ordering is maintained

Given two riffle independent subsets

Set A={1, 2, 3} and $p_A(\sigma_A)$
Set B={4, 5, 6} and $p_B(\sigma_B)$

Ranking process of a riffle independent set

$shuffle(\sigma_A, \sigma_B) = shuffle((213),(465)) = [(421635),(421653),\ldots]$

UPV - EHU

Distribution over the shuffling, $p(shuffle(A, B))$

$$p((RLLRLR)) = 0.1$$
$$p((RLLRRL)) = 0.12;$$
$$\vdots$$

Definition

Distribution p is said to be riffle independent if
$p(\sigma) = p(shuffle(A, B)) \cdot p_A(\sigma_A) \cdot p_B(\sigma_B)$

Simplifications

- Different settings of the interleaving distribution lead to simpler models
- A generalization of the the full independent model

To take home

- Natural criterion in the ranking domain
- First generate the rankings on the independent subsets and then shuffle those rankings to obtain the complete one

Summary

- L-decomposability: the probability of selecting an item at each stage does not depend on the order of the already selected items
- First order independence: a subset of items A always map into a subset of the positions X ($|A| = |X|$)
- Riffle independence: given 2 riffle independent subsets A and B and items $a_1, a_2 \in A$ and $b \in B$ knowing that a_1 is ranked before a_2 does not give any info about where b is ranked

UPV - EHU

Outline of the talk

UPV - EHU

- Application of ranking has changed in history, from psychology to machine learning
- So has the size and amount of data, from small datasets to huge amount of data

Kinds of data

- Partial ranking: not every items has received a rank, top 5 ranking
- Rank with ties: a set of items is preferred to another set but there is not order within the items in the set
- Implicit data: the acts of users are known (buying film A or visiting B web page) but not their opinion about it, there is no rating or ranking no negative feedback

German sample (Croon, 1988)

Small database of the permutations of 4 elements cited in (Bückenholt, U. 2002)

Idea (Fligner et al. 1986)

98 college students where asked to rank five words, namely thought, play, theory, dream, and attention regarding its association with the word idea
Paper that cite it are (Gupta et al. 2002)

Cars (Bückenholt, 2002)

279 Spanish college students were asked to rank four compact cars according to their purchase preferences. The authors provide the ranking patterns' observed frequencies in this sample

APA (Diaconis, 1988)

- The American Psychological Association dataset includes 15449 ballots of the election of the president in 1980, 5738 of which are complete rankings, in which the candidates are ranked from most to least favorite
- Candidates belong to 3 different schools: clinical, research and community
- It is very frequently used (Huang et. al 2010)

Sushi (Kamishima et al. 2010)

This information collected via web comprises user information as well as their preferences in sushi, including

- 1025 partial rankings on 100 different kinds of sushi
- 1025 full rankings on 10 different kinds of sushi (items are ranked by every ranker)

Cited by (Huang et. al 2010)

Irish Election (Gormley, et al. (2006))

Results of the election for the Irish House of Parliament election dataset from the Meath constituency in Ireland. It used also in (Huang et. al 2010)

Paired comparisons (Hunter, 2003)

- NFL dataset, results of the 1997 NFL season
- NASCAR dataset, results of the 2002 car season, 36 races and 83 divers took part

Jester (Goldberg et al. 2001)

- 4.1 Million continuous ratings (-10.00 to +10.00) of 100 jokes from 73421 users
- Appears in (Meila et al. 2010)

Both are offline

Netflix (Bennett, et al. 07)

Movie recomender system. In 2009 1000000 dollar prize was given to Bellkor's programatic chaos for besting Netflix's recomender
offline for privacy reasons

WikiLens

WikiLens was a generalized collaborative recommender system that allowed its community to define item types (movie, book album, restaurant, web) and categories for each type, and then rate and get recommendations for items

Book crossing (Ziegler, 2005)

Collected from the Book-Crossing community.
Contains 278.858 users (with demographic information) providing 1.149.780 ratings (explicit / implicit) about 271.379 books

MovieLens

Rankings with-ties

- MovieLens 100k: 100000 ratings (1-5) from 1000 users on 1700 movies.
- MovieLens 1M: 1 million ratings from 6000 users on 4000 movies.
- MovieLens 10M: 10 million ratings and 100000 tag applications applied to 10000 movies by 72000 users. In the dataset, the movies are linked to movie review systems. Each movie does have its IMDb and RT identifiers, English and Spanish titles, picture URLs, genres, directors, actors (ordered by ”popularity”), RT audience’ and experts’ ratings and scores, countries, and filming locations.

HetRec 2011 conference

Highly sparse databases

- movielens 2k: extension of MovieLens10M dataset, which contains personal ratings and tags about movies.
- delicious 2k: implicit data from Delicious social bookmarking system. Each of the 1867 users has bookmarks, tag assignments, i.e. tuples [user, tag, bookmark], and contact relations within the dataset social network. Each bookmark has a title and URL.
- lastfm 2k: implicit dataset in which each user has a list of most listened music artists, tag assignments, i.e. tuples [user, tag, artist], and friend relations within the dataset social network. Each artist has a Last.fm URL and a picture URL.

Non parametric models

- The SnOB software provides general functions or permutations such as Fast Fourier transform
- The props toolbox (based on SnOB) provides some efficient inference algorithms for distributions over permutations and examples as described in the paper (Huang et al. 2009)

UPV - EHU

Mallows model

pyMallows is a set of Python routines for fitting and simulating from a generalized Mallows model based on Kendall's-τ distance. Learning algorithms implemented for both full and partial rankings.

Paired comparisons

- *prefmod* fits and simulates data as pairs comparisons, Bradley-Terry models and others, and their extension. The package includes datasets for testing
- *BradleyTerry2, eba, psychotree* are other R packages for paired comparisons
- *MMBT* is a Matlab packages for the estimation of the Bradley-Terry model and and its extension to multi comparison case

UPV - EHU

Outline of the talk

Conclusions

- Many process are based on orderings and ranking
- New technologies allow to store thousand partial rankings of thousand of items
- New probabilistic approaches need to be developed to deal with them
- Two ways:
  - To adapt well-known probabilistic models
  - To create new specific models for the new kind of data

UPV - EHU