Clustering Rankings in the Fourier Domain

Stéphan Clémençon and **Romaric Gaudel** and **Jérémie Jakubowicz**

LTCI, Telecom Paristech (TSI)
UMR Institut Telecom/CNRS No. 5141

ECML PKDD, September 2011

Distributions on rankings

- Many applications consider ranked data / distributions on rankings

(Uniform distribution with respect to constraints)

  - Top-k lists
    - ⋆ Rank of the k most preferred objects
      $3 > 2 > 5 > \dots$

amazon.com ebay millemercis.com

  - Preference data
    - ⋆ Preferences on k (randomly) picked objects
      $\dots > 3 > \dots > 2 > \dots > 5 > \dots$

“sushi” dataset

  - Bucket order
    - ⋆ Preferences on groups of objects
      $3, 2 > 5, 1, 7 > 4, 6, 8$

NETFLIX

TELECOM ParisTech cnrs

Representation for distributions on Rankings

- Probability table
  - $n!$ (factorial n) coefficients
- Fourier representation [Diaconis, 1989; Kondor & Barbosa, 2010]
  - $n!$ coefficients
  - Few relevant coefficients in practice
- Parametric models
  - Mallows [Mallows, 1957]
  - Plackett-Luce [Luce, 1959; Plackett, 1975]

Representation for distributions on Rankings

- Probability table
  - $n!$ (factorial n) coefficients
- Fourier representation [Diaconis, 1989; Kondor & Barbosa, 2010]
  - $n!$ coefficients
  - Few relevant coefficients in practice
- Parametric models
  - Mallows [Mallows, 1957]
  - Plackett-Luce [Luce, 1959; Plackett, 1975]

Contributions

Clustering of rankings through sparse Fourier representation

- Position
  - Clustering of distributions on rankings
    - ⋆ Gather ranking distributions with similar shapes
- Proposed approach
  - Work in the Fourier representation

$\Longrightarrow$
    - ⋆ Sparse representation of 1 distribution
    - ⋆ Sparse difference between representations of 2 distributions

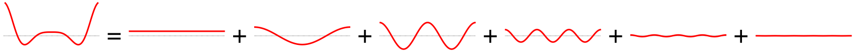
Outline

Fourier representation

For real line function

- Functions are decomposed on the sinusoidal basis

$$f(x) = 1.1 + 2.1\cos(x) + 3.2\cos(2x) + 1.5\cos(3x) + 0.2\cos(4x) + 0.01\cos(5x) + \ldots$$



- The information is contained in few (low frequency) coefficients

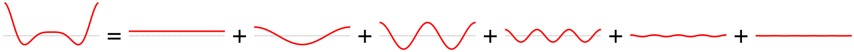
  $\Longrightarrow$ Reduced storage/transfer/computation costs

Fourier representation

For real line function

- Functions are decomposed on the sinusoidal basis

$$f(x) = 1.1 + 2.1\cos(x) + 3.2\cos(2x) + 1.5\cos(3x) + 0.2\cos(4x) + 0.01\cos(5x) + \ldots$$



- The information is contained in few (low frequency) coefficients

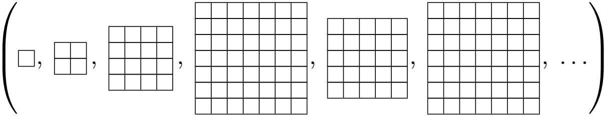
$\Longrightarrow$ Reduced storage/transfer/computation costs

Fourier representation

For functions on $\mathfrak{S}_n$ [Diaconis, 1989]

- There is no simple basis (corresponding to eigen-spaces of dimension 1)
  $\Longrightarrow$ Fourier coefficients are matrices indexed by the set $\mathcal{R}_n$ of all *integer partitions of n*

$\mathcal{F}f =$

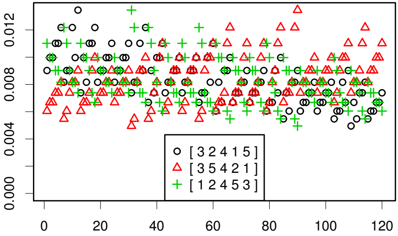


$$\mathcal{R}_n = \left\{ \left\{ \xi = (n_1, \ldots, n_k) \in \mathbb{N}^{*k} : \; n_1 \geq \cdots \geq n_k, \sum_{i=1}^{k} n_i = n \right\}, 1 \leq k \leq n \right\}$$

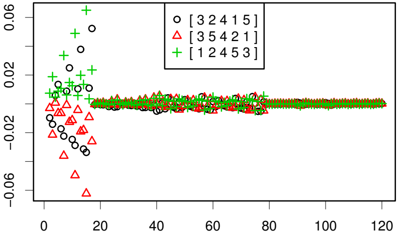
- “Low-frequency” coefficients are related to low order summaries ($\mathbb{P}[\sigma(i,j) = (k,\ell)]$)

Example: Mallows($\mathfrak{S}_5$)

Exponential distribution on rankings, $\gamma = 0.1$



"Temporal" coefficents



Fourier coefficients

- Remark:
  - A few relevant parameters when using the Fourier representation

Uncertainty principle

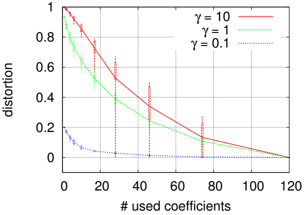
Balancing Sparsity

Theorem (inspired from [Donoho & Stark, 1989])

Let $f \in \mathbb{C}[\mathbb{S}_n]$ of Fourier transform $\mathcal{F}f$. Denote by $\mathrm{supp}(f) = \{\sigma \in \mathbb{S}_n : \ f(\sigma) \neq 0\}$ and by $\mathrm{supp}(\mathcal{F}f) = \{\xi \in \mathcal{R}_n : \ \mathcal{F}f(\xi) \neq 0\}$ the support of f and that of its Fourier transform respectively. Then, we have:

$$\#\mathrm{supp}(f) \cdot \sum_{\xi \in \mathrm{supp}(\mathcal{F}f)} d_\xi^2 \geq n!.$$

- Direct consequence
  - Both representations cannot be simultaneously sparse



Distortion with Mallows($\mathfrak{S}_5$)

Outline

Clustering of rankings

- Aim
  - Gather distributions on rankings with similar shape
- Objective function
  - Minimize (on all partitions $\mathcal{C}$)

$$
\begin{aligned}
\widehat{\mathcal{M}}(\mathcal{C}) &= \sum_{l=1}^{L} \sum_{1 \leq i,\, j \leq N} \|f_i - f_j\|^2 \cdot \mathbb{I}\{(f_i, f_j) \in \mathcal{C}_l^2\} \\
&= \frac{1}{n!} \sum_{\xi \in \mathcal{R}_n} d_\xi \sum_{l=1}^{L} \sum_{1 \leq i,\, j \leq N:\ (f_i, f_j) \in \mathcal{C}_l^2} \|\mathcal{F}f_i(\xi) - \mathcal{F}f_j(\xi)\|^2_{HS(d_\xi)}
\end{aligned}
$$

with $d_\xi \times d_\xi$ the dimension of the matrix indexed by ξ

Managing sparsity

- Aim
  - Gather distributions on rankings with similar shape
  - Use few Fourier coefficients
- New objective function [Witten & Tibshirani, 2010]
  - Minimize (on all partitions $\mathcal{C}$, and all weight vectors ω)

$$\widehat{\mathcal{M}}_\omega(\mathcal{C}) = \sum_{\xi \in \mathcal{R}_n} \omega_\xi \frac{d_\xi}{n!} \sum_{l=1}^{L} \sum_{1 \leq i,\, j \leq N:\, (f_i, f_j) \in \mathcal{C}_l^2} ||\mathcal{F} f_i(\xi) - \mathcal{F} f_j(\xi)||^2_{HS(d_\xi)}$$

with $\omega = (\omega_\xi)_{\xi \in \mathbb{R}_n} \in \mathbb{R}_+^{\#\mathcal{R}_n}$, $||\omega||^2_{l_2} \leq 1$ and $||\omega||_{l_1} \leq \lambda$

- Remark:
  - Fixing $\omega = (1/\sqrt{\#\mathcal{R}_n}, \ldots, 1/\sqrt{\#\mathcal{R}_n})$ leads to the initial optimization problem (without ω)

Algorithm

- Initialize $\omega = (1/\sqrt{\#\mathcal{R}_n}, \ldots, 1/\sqrt{\#\mathcal{R}_n})$
- Until convergence, iterate steps 1 and 2
  1. Fixing the weight vector ω, minimize $\widehat{\mathcal{M}}_\omega(\mathcal{C})$ after the partition $\mathcal{C}$
  2. Fixing the partition $\mathcal{C}$, minimize $\widehat{\mathcal{M}}_\omega(\mathcal{C})$ after ω.
- Remarks
  - Step 1 is performed by a standard clustering algorithm
  - Step 2 accepts a closed form [Witten & Tibshirani, 2010]

Outline

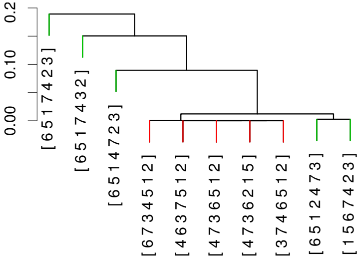
1. Sparsity in the Fourier Representation
2. Sparse Clustering of Rankings
3. Numerical Experiments

Experiments

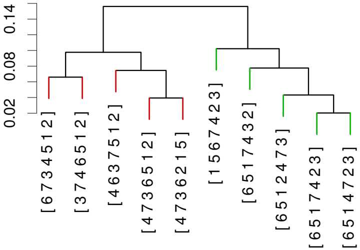
- Aim
  - Recover clustering information
  - Use few coefficients
- Datasets
  - Mallows (synthetic)
    - Exponential distribution on rankings
  - Top-k lists (synthetic)
    - Uniform distribution on rankings
  - E-commerce Dataset
    - List of purchased products (ordered by date)

Mallows($\mathfrak{S}_7$)

$\gamma = 1$



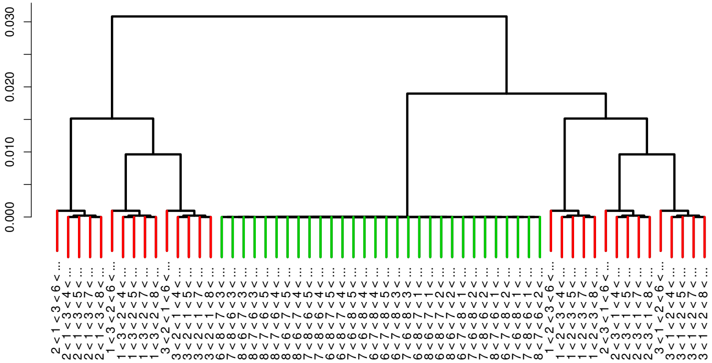
"Temporal" representation
(3 coefficients selected)



Fourier representation
(54 coefficients selected)

- Remarks:
  - The Fourier representation recovers the clustering information
  - The Fourier representation uses few coefficients (compared to $n! = 5,040$)

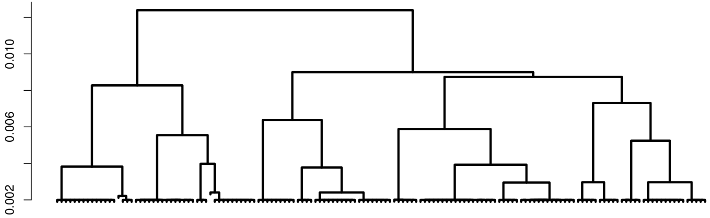
Top-4 lists on $\mathfrak{S}_8$



Fourier representation (7 coefficients selected)

- Remarks:
  - The Fourier representation recovers the clustering information
  - The “temporal” representation is useless (examples have disjoint supports)
  - The Fourier representation uses few coefficients

E-commerce dataset



Fourier representation (5 coefficients selected)

- Remarks:
  - 4 groups among users
  - Focuses on few coefficients

Conclusion and perspectives

Conclusion

- A new approach to cluster rankings
  - Based on the Fourier representation
    $\Longrightarrow$ Sparse representation
  - Based on a sparse clustering criterion
    $\Longrightarrow$ Focuses on relevant coefficients
- Several theoretical and numerical results supporting the approach

- Future work
  - Better understanding of the class of distributions with sparse Fourier representation

Thank you

Bibliography

P. Diaconis *A generalization of spectral analysis with application to ranked data*. Ann. Stat. 89
D. Donoho and P. Stark *Uncertainty Principles and Signal Recovery*. SIAM J. Appl. Math. 89
R. Kondor and M. Barbosa *Ranking with kernels in Fourier space*. COLT'10
R. D. Luce *Individual Choice Behavior*. Wiley 59
C. L. Mallows *Non-Null Ranking Models*. Biometrika 57
R. L. Plackett *The analysis of permutations*. Appl. Stat. 75
D. Witten *A Framework for Feature Selection in Clustering*. JASA'10