A Learning Theory of Ranking Aggregation

France/Japan Machine Learning Workshop

Anna Korba, Stephan Clémençon, Eric Sibony

November 14, 2017

Télécom ParisTech

Outline

1. The Ranking Aggregation Problem
2. Statistical Framework
3. Results

The Ranking Aggregation Problem

The Ranking Aggregation Problem

Framework

- n items: $\{1, \ldots, n\}$.
- N rankings on the n items (from most preferred to last): $i_1 \succ i_2 \succ \cdots \succ i_n$.
- $i_1 \succ \cdots \succ i_n \Longleftrightarrow$ permutation σ on $\{1, \ldots, n\}$ s.t. $\sigma(i_j) = j$.

The Ranking Aggregation Problem

Framework

- n items: $\{1, \ldots, n\}$.
- N rankings on the n items (from most preferred to last): $i_1 \succ i_2 \succ \cdots \succ i_n$.
- $i_1 \succ \cdots \succ i_n \iff$ permutation σ on $\{1, \ldots, n\}$ s.t. $\sigma(i_j) = j$.

The Ranking Aggregation Problem

Framework

- n items: $\{1, \ldots, n\}$.
- N rankings on the n items (from most preferred to last): $i_1 \succ i_2 \succ \cdots \succ i_n$.
- $i_1 \succ \cdots \succ i_n \iff$ permutation σ on $\{1, \ldots, n\}$ s.t. $\sigma(i_j) = j$.

Suppose we have a dataset of rankings/permutations $(\sigma_1, \ldots, \sigma_N) \in \mathfrak{S}_n^N$.

The Ranking Aggregation Problem

Framework

- n items: $\{1, \ldots, n\}$.
- N rankings on the n items (from most preferred to last): $i_1 \succ i_2 \succ \cdots \succ i_n$.
- $i_1 \succ \cdots \succ i_n \iff$ permutation σ on $\{1, \ldots, n\}$ s.t. $\sigma(i_j) = j$.

Suppose we have a dataset of rankings/permutations $(\sigma_1, \ldots, \sigma_N) \in \mathfrak{S}_n^N$.

Consensus Ranking

We want to find a global order ("consensus") σ^* on the n items that best represents the dataset.

Copeland Rule.

Sort the items according to their Copeland score, defined for each item i by:

$$s_C(i) = \frac{1}{N} \sum_{t=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{I}[\sigma_t(i) < \sigma_t(j)]$$

Copeland Rule.

Sort the items according to their Copeland score, defined for each item i by:

$$s_C(i) = \frac{1}{N} \sum_{t=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{I}[\sigma_t(i) < \sigma_t(j)]$$

Kemeny's rule (1959).

Find the solution of :

$$min_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^{N} d(\sigma, \sigma_t) \tag{1}$$

where d is the Kendall's tau distance:

$$d_\tau(\sigma, \sigma') = \sum_{i<j} \mathbb{I}\{(\sigma(i) - \sigma(j))(\sigma'(i) - \sigma'(j)) < 0\},$$

Kemeny's consensus has a lot of interesting properties, but it is NP-hard to compute, even for $n = 4$ (see Dwork et al., 2001).

Statistical Framework

Statistical Reformulation

Suppose the dataset is composed of N i.i.d. copies $\Sigma_1, \dots, \Sigma_N$ of a r.v. $\Sigma \sim P$. A (Kemeny) **median** of P w.r.t. d is solution of:

$$\boxed{\min_{\sigma \in \mathfrak{S}_n} \mathbb{E}_{\Sigma \sim P}[d(\Sigma, \sigma)],}$$

where $L(\sigma) = \mathbb{E}_{\Sigma \sim P}[d(\Sigma, \sigma)]$ is **the risk** of σ.

Statistical Reformulation

Suppose the dataset is composed of N i.i.d. copies $\Sigma_1, \ldots, \Sigma_N$ of a r.v. $\Sigma \sim P$. A (Kemeny) **median** of P w.r.t. d is solution of:

$$\min_{\sigma \in \mathfrak{S}_n} \mathbb{E}_{\Sigma \sim P}[d(\Sigma, \sigma)],$$

where $L(\sigma) = \mathbb{E}_{\Sigma \sim P}[d(\Sigma, \sigma)]$ is **the risk** of σ.

Let $\widehat{L}_N(\sigma) = \frac{1}{N} \sum_{t=1}^{N} d(\Sigma_t, \sigma)$.

Statistical Reformulation

Suppose the dataset is composed of N i.i.d. copies $\Sigma_1, \ldots, \Sigma_N$ of a r.v. $\Sigma \sim P$. A (Kemeny) **median** of P w.r.t. d is solution of:

$$\min_{\sigma \in \mathfrak{S}_n} \mathbb{E}_{\Sigma \sim P}[d(\Sigma, \sigma)],$$

where $L(\sigma) = \mathbb{E}_{\Sigma \sim P}[d(\Sigma, \sigma)]$ is **the risk** of σ.

Let $\widehat{L}_N(\sigma) = \frac{1}{N} \sum_{t=1}^{N} d(\Sigma_t, \sigma)$.

Goal of our analysis:

Study the performance of **Kemeny empirical medians**, i.e. solutions $\widehat{\sigma}_N$ of:

$$\min_{\sigma \in \mathfrak{S}_n} \widehat{L}_N(\sigma),$$

through the excess of risk $L(\widehat{\sigma}_N) - L^*$.

$\Rightarrow$ We establish links with Copeland method.

Risk of Ranking Aggregation

The risk of a median σ is $L(\sigma) = \mathbb{E}_{\Sigma \sim P}[d(\Sigma, \sigma)]$, where d is the Kendall's tau distance:

$$d(\sigma, \sigma') = \sum_{\{i,j\} \subset [\![n]\!]} \{(\sigma(i) - \sigma(j))(\sigma'(i) - \sigma'(j)) < 0\}$$

Let $p_{i,j} = \mathbb{P}[\Sigma(i) < \Sigma(j)]$ the probability that item i is preferred to item j.

Risk of Ranking Aggregation

The risk of a median σ is $L(\sigma) = \mathbb{E}_{\Sigma \sim P}[d(\Sigma, \sigma)]$, where d is the Kendall's tau distance:

$$d(\sigma, \sigma') = \sum_{\{i,j\} \subset [\![n]\!]} \{(\sigma(i) - \sigma(j))(\sigma'(i) - \sigma'(j)) < 0\}$$

Let $p_{i,j} = \mathbb{P}[\Sigma(i) < \Sigma(j)]$ the probability that item i is preferred to item j.

The risk can be rewritten:

$$L(\sigma) \; = \; \sum_{i<j} p_{i,j} \mathbb{I}\{\sigma(i) \; > \; \sigma(j)\} + \sum_{i<j} (1 \, - \, p_{i,j}) \mathbb{I}\{\sigma(i) \; < \; \sigma(j)\}.$$

Risk of Ranking Aggregation

The risk of a median σ is $L(\sigma) = \mathbb{E}_{\Sigma \sim P}[d(\Sigma, \sigma)]$, where d is the Kendall's tau distance:

$$d(\sigma, \sigma') = \sum_{\{i,j\} \subset [\![n]\!]} \{(\sigma(i) - \sigma(j))(\sigma'(i) - \sigma'(j)) < 0\}$$

Let $p_{i,j} = \mathbb{P}[\Sigma(i) < \Sigma(j)]$ the probability that item i is preferred to item j.

The risk can be rewritten:

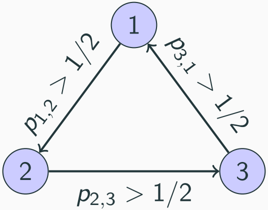
$$L(\sigma) \;=\; \sum_{i<j} p_{i,j} \mathbb{I}\{\sigma(i) \;>\; \sigma(j)\} + \sum_{i<j} (1 \;-\; p_{i,j}) \mathbb{I}\{\sigma(i) \;<\; \sigma(j)\}.$$

So if there exists a permutation σ verifying: $\forall i < j$ s.t. $p_{i,j} \neq 1/2$,

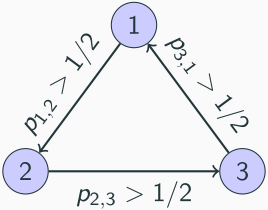
$$(\sigma(j) - \sigma(i)) \cdot (p_{i,j} - 1/2) > 0, \tag{2}$$

it would be necessary of median for P.

Preference cycles

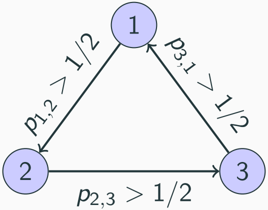


Preference cycles



$\rightarrow$ No permutation can satisfy this condition for each pair of items!

Preference cycles



$\rightarrow$ No permutation can satisfy this condition for each pair of items!

Definition

P on $\mathfrak{S}_n$ is **stochastically transitive** if : $\forall (i, j, k) \in [\![n]\!]^3$,

$$p_{i,j} \geq 1/2 \text{ and } p_{j,k} \geq 1/2 \Rightarrow p_{i,k} \geq 1/2.$$

Moreover, if $p_{i,j} \neq 1/2$ for all $i < j$, P is **strictly stochastically transitive**.

Results

Optimality

Theorem

- If P is stochastically transitive, there exists $\sigma^* \in \mathfrak{S}_n$ verifying: $\forall i < j$ s.t. $p_{i,j} \neq 1/2$,

$$(\sigma(j) - \sigma(i)) \cdot (p_{i,j} - 1/2) > 0,$$

is verified.

- The **Copeland score** of an item i, that is:

$$\boxed{s^*(i) = 1 + \sum_{k \neq i} \mathbb{I}\{p_{i,k} < \frac{1}{2}\}}$$

defines a permutation $s^* \in \mathfrak{S}_n$ and is the unique median of P iff P is strictly stochastically transitive.

Optimality

Theorem

- If P is stochastically transitive, there exists $\sigma^* \in \mathfrak{S}_n$ verifying: $\forall i < j$ s.t. $p_{i,j} \neq 1/2$,

$$(\sigma(j) - \sigma(i)) \cdot (p_{i,j} - 1/2) > 0,$$

  is verified.

- The **Copeland score** of an item i, that is:

$$\boxed{s^*(i) = 1 + \sum_{k \neq i} \mathbb{I}\{p_{i,k} < \frac{1}{2}\}}$$

  defines a permutation $s^* \in \mathfrak{S}_n$ and is the unique median of P iff P is strictly stochastically transitive.

Universal Rates

The excess risk of $\widehat{\sigma}_N$ is upper bounded:

(i) In expectation by

$$\mathbb{E}\left[L(\widehat{\sigma}_N) - L^*\right] \leq \frac{n(n-1)}{2\sqrt{N}}$$

(ii) With probability higher than $1-\delta$ for any $\delta \in (0,1)$ by

$$L(\widehat{\sigma}_N) - L^* \leq \frac{n(n-1)}{2}\sqrt{\frac{2\log(n(n-1)/\delta)}{N}}.$$

Conditions for Fast Rates

Suppose that P verifies:

- the Stochastic Transitivity condition:

$$p_{i,j} \geq 1/2 \text{ and } p_{j,k} \geq 1/2 \Rightarrow p_{i,k} \geq 1/2.$$

- the Low-Noise condition **NA**(h) for some $h > 0$:

$$\min_{i<j} |p_{i,j} - 1/2| \geq h.$$

Conditions for Fast Rates

Suppose that P verifies:

- the Stochastic Transitivity condition:

$$p_{i,j} \geq 1/2 \text{ and } p_{j,k} \geq 1/2 \Rightarrow p_{i,k} \geq 1/2.$$

- the Low-Noise condition **NA**(h) for some $h > 0$:

$$\min_{i<j} |p_{i,j} - 1/2| \geq h.$$

$\Rightarrow$ Introduced for binary classification (see Koltchinskii and Beznosova, 2005).
$\Rightarrow$ Used for estimation of matrix of pairwise probabilities (see Shah et al., 2016).

Fast rates

Let $\alpha_h = \frac{1}{2}\log\left(1/(1-4h^2)\right)$.

Assume that P satisfies the previous conditions.

(i) For any empirical Kemeny median $\widehat{\sigma}_N$, we have:

$$\mathbb{E}\left[L(\widehat{\sigma}_N) - L^*\right] \leq \frac{n^2(n-1)^2}{8} e^{-\alpha_h N}.$$

(ii) With probability at least $1 - (n(n-1)/4)e^{-\alpha_h N}$, the empirical Copeland score

$$\widehat{s}_N(i) = 1 + \sum_{k \neq i} \mathbb{I}\{\widehat{p}_{i,k} < \frac{1}{2}\}$$

for $1 \leq i \leq n$ belongs to $\mathfrak{S}_n$ and is the unique solution of Kemeny empirical minimization.

$\Rightarrow$ In practice: under the needed conditions, Copeland method ($\mathcal{O}(N\binom{n}{2})$) outputs the Kemeny consensus (NP-hard) with high prob.

Conclusion and future directions

- We introduced a general statistical framework for ranking aggregation and established rates of convergence.

Conclusion and future directions

- We introduced a general statistical framework for ranking aggregation and established rates of convergence.
- We can write the empirical risk when one observe pairwise comparisons only instead of full rankings

Conclusion and future directions

- We introduced a general statistical framework for ranking aggregation and established rates of convergence.
- We can write the empirical risk when one observe pairwise comparisons only instead of full rankings
- Conditions for fast rates have good chance to be satisfied in a homogeneous population so Copeland method could be useful for local methods

Conclusion and future directions

- We introduced a general statistical framework for ranking aggregation and established rates of convergence.
- We can write the empirical risk when one observe pairwise comparisons only instead of full rankings
- Conditions for fast rates have good chance to be satisfied in a homogeneous population so Copeland method could be useful for local methods