Inferring from an imprecise Plackett–Luce model: application to label ranking

SUM 2020

Loïc Adam, Arthur Van Camp, Sébastien Destercke & Benjamin Quost

Laboratoire Heudiasyc - Université de Technologie de Compiègne

19th September 2020

Label ranking

Consider a finite set of labels $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$.

Label ranking

Consider a finite set of labels $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$.

We want to rank them by a complete, transitive and asymmetric relation $\succ$.

Label ranking

Consider a finite set of labels $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$.

We want to rank them by a complete, transitive and asymmetric relation $\succ$.

Central question: How can we model beliefs about $\succ$?

Label ranking

Consider a finite set of labels $\Lambda = \{\lambda_1, \dots, \lambda_n\}$.

We want to rank them by a complete, transitive and asymmetric relation $\succ$.

Central question: How can we model beliefs about $\succ$?

$\succ$ characterises a unique permutation τ (called label ranking) of $\{1, \dots, n\}$:

$$\lambda_{\tau(1)} \succ \lambda_{\tau(2)} \succ \cdots \succ \lambda_{\tau(n)}.$$

Label ranking

Consider a finite set of labels $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$.

We want to rank them by a complete, transitive and asymmetric relation $\succ$.

Central question: How can we model beliefs about $\succ$?

$\succ$ characterises a unique permutation τ (called label ranking) of $\{1, \ldots, n\}$:

$$\lambda_{\tau(1)} \succ \lambda_{\tau(2)} \succ \cdots \succ \lambda_{\tau(n)}.$$

A popular belief model on the set $\mathcal{L}(\Lambda)$ of all label rankings is the Plackett–Luce (PL) model

$$P_v(\tau) := \prod_{k=1}^{n} \frac{v_{\tau(k)}}{\sum_{\ell=k}^{n} v_{\tau(\ell)}}$$

with positive strength vector $v = (v_1, \ldots, v_n)$. We assume $\sum_{k=1}^{n} v_k = 1$.

Plackett–Luce (PL) model

The Plackett–Luce (PL) model

$$P_v(\tau) := \prod_{k=1}^{n} \frac{v_{\tau(k)}}{\sum_{\ell=k}^{n} v_{\tau(\ell)}} \quad \text{with} \sum_{k=1}^{n} v_k = 1.$$

How can we find the best (most probable) ranking $\tau^\star$?

Plackett–Luce (PL) model

The Plackett–Luce (PL) model

$$P_v(\tau) := \prod_{k=1}^{n} \frac{v_{\tau(k)}}{\sum_{\ell=k}^{n} v_{\tau(\ell)}} \quad \text{with} \sum_{k=1}^{n} v_k = 1.$$

How can we find the best (most probable) ranking $\tau^\star$?

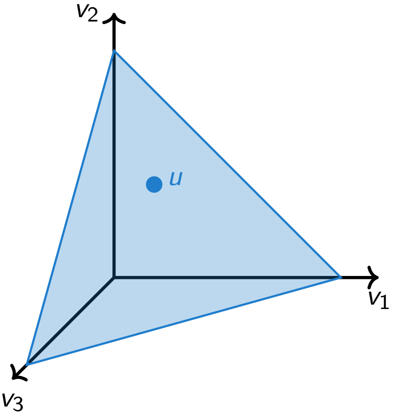
$$\tau^\star \text{ maximises } P_v(\cdot) \Leftrightarrow v_{\tau^*(1)} \geq v_{\tau^*(2)} \geq v_{\tau^*(3)} \cdots \geq v_{\tau^*(n-1)} \geq v_{\tau^*(n)}.$$

Imprecise Plackett–Luce (IPL) model

The Plackett–Luce (PL) model: $P_v(\tau) := \prod_{k=1}^{n} \frac{v_{\tau(k)}}{\sum_{\ell=k}^{n} v_{\tau(\ell)}}$.

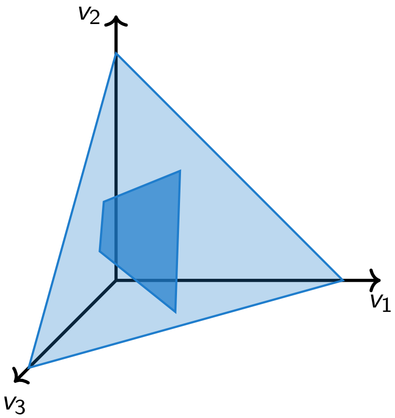
Imprecise Plackett–Luce (IPL) model

The Plackett–Luce (PL) model: $P_v(\tau) := \prod_{k=1}^{n} \frac{v_{\tau(k)}}{\sum_{\ell=k}^{n} v_{\tau(\ell)}}$.



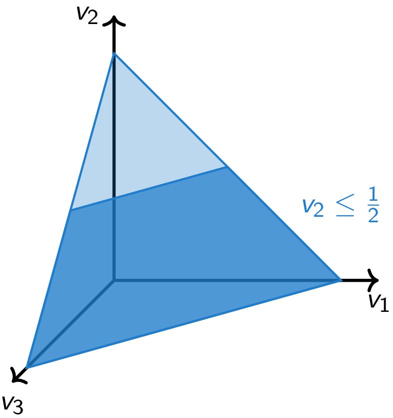
Imprecise Plackett–Luce (IPL) model

The Plackett–Luce (PL) model: $P_v(\tau) := \prod_{k=1}^{n} \frac{v_{\tau(k)}}{\sum_{\ell=k}^{n} v_{\tau(\ell)}}$.



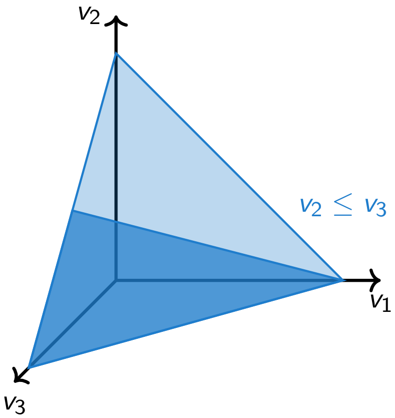
Imprecise Plackett–Luce (IPL) model

The Plackett–Luce (PL) model: $P_v(\tau) := \prod_{k=1}^{n} \frac{v_{\tau(k)}}{\sum_{\ell=k}^{n} v_{\tau(\ell)}}$.



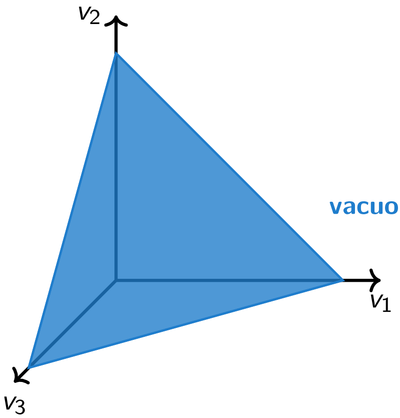
Imprecise Plackett–Luce (IPL) model

The Plackett–Luce (PL) model: $P_v(\tau) := \prod_{k=1}^{n} \frac{v_{\tau(k)}}{\sum_{\ell=k}^{n} v_{\tau(\ell)}}$.



Imprecise Plackett–Luce (IPL) model

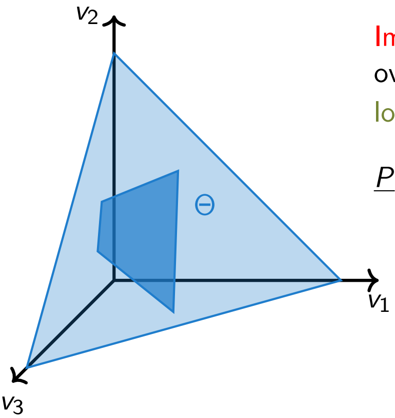
The Plackett–Luce (PL) model: $P_v(\tau) := \prod_{k=1}^{n} \frac{v_{\tau(k)}}{\sum_{\ell=k}^{n} v_{\tau(\ell)}}$.



vacuous belief model

Imprecise Plackett–Luce (IPL) model

The Plackett–Luce (PL) model: $P_v(\tau) := \prod_{k=1}^{n} \frac{v_{\tau(k)}}{\sum_{\ell=k}^{n} v_{\tau(\ell)}}$.



Imprecise Plackett–Luce model: let v vary over a set Θ.

lower and upper probability of τ:

$$\underline{P}_\Theta(\tau) := \inf_{v \in \Theta} P_v(\tau); \quad \overline{P}_\Theta(\tau) := \sup_{v \in \Theta} P_v(\tau).$$

Decision-making with IPL

Given an IPL with set of strength vectors Θ, how can we find the "best" or "most probable" rankings?

Problem: there is in general no unique best ranking.

Decision-making with IPL

Given an IPL with set of strength vectors Θ, how can we find the "best" or "most probable" rankings?

Problem: there is in general no unique best ranking.

Let $\tau_1 \succ \tau_2$ when $P_v(\tau_1) > P_v(\tau_2)$ for all v in Θ.

Decision-making with IPL

Given an IPL with set of strength vectors Θ, how can we find the "best" or "most probable" rankings?

Problem: there is in general no unique best ranking.

Let $\tau_1 \succ \tau_2$ when $P_v(\tau_1) > P_v(\tau_2)$ for all v in Θ. The set of undominated rankings

$$\mathcal{M}_\Theta = \{\tau \in \mathcal{L}(\Lambda) : (\forall \tau' \in \mathcal{L}(\Lambda))\tau' \not\succ \tau\}$$

contains the (Walley–Sen) maximal rankings.

Decision-making with IPL

Given an IPL with set of strength vectors Θ, how can we find the "best" or "most probable" rankings?

Problem: there is in general no unique best ranking.

Let $\tau_1 \succ \tau_2$ when $P_v(\tau_1) > P_v(\tau_2)$ for all v in Θ. The set of undominated rankings

$$\mathcal{M}_\Theta = \{\tau \in \mathcal{L}(\Lambda) : (\forall \tau' \in \mathcal{L}(\Lambda))\tau' \not\succ \tau\}$$

contains the (Walley–Sen) maximal rankings.

As another approach, the set

$$\mathcal{E}_\Theta = \bigcup_{v \in \Theta} \{\tau \text{ maximises } P_v(\cdot)\}$$

contains the E-admissible rankings.

Decision-making with IPL

Given an IPL with set of strength vectors Θ, how can we find the "best" or "most probable" rankings?

Problem: there is in general no unique best ranking.

Let $\tau_1 \succ \tau_2$ when $P_v(\tau_1) > P_v(\tau_2)$ for all v in Θ. The set of undominated rankings

$$\mathcal{M}_\Theta = \{\tau \in \mathcal{L}(\Lambda) : (\forall \tau' \in \mathcal{L}(\Lambda))\tau' \not\succ \tau\}$$

contains the (Walley–Sen) maximal rankings.

As another approach, the set

$$\mathcal{E}_\Theta = \bigcup_{v \in \Theta} \{\tau \text{ maximises } P_v(\cdot)\}$$

contains the E-admissible rankings. We always have $\mathcal{E}_\Theta \subseteq \mathcal{M}_\Theta$.

Inferences with an IPL

We establish two theoretical results.

Inferences with an IPL

We establish two theoretical results.

1. For arbitrary sets Θ, we find an efficient algorithm to calculate an outer approximation of the maximal rankings $\mathcal{M}_\Theta$.

Inferences with an IPL

We establish two theoretical results.

1. For arbitrary sets Θ, we find an efficient algorithm to calculate an outer approximation of the maximal rankings $\mathcal{M}_\Theta$.
2. For a specific type of Θ—namely, when it is generated by intervals—we find two exact algorithms:
   1. an efficient algorithm to determine whether a given ranking is E-admissible;
   2. another algorithm to calculate all the E-admissible rankings $\mathcal{E}_\Theta$.

Problem example

Individual	Preferred object	Less preferred object	Disliked object
John	Cheese	Meat	Parrot
Michael	Parrot	Cheese	Meat
Terry	Meat	Cheese	Parrot

Table: Rankings of objects

Individual	Temperament	Self-esteem
John	Nervous	Too much
Michael	Calm	Too much
Terry	Nervous	Perhaps

Table: Characteristics of individuals

New individual Graham : calm and not too much self-esteem $\Rightarrow$

$$\text{Parrot} \succ \text{Meat} \succ \text{Cheese}.$$

Preference learning and motivation for a robust approach

- Mapping between an instance $\mathbf{x} \in \mathcal{X}$ and a order of labels $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$ called ranking. Rankings can be incomplete.
- The probability of each $\mathcal{L}(\Lambda)$ is here modelled by a Plackett–Luce model. τ^* is the permutation with the highest probability.

- In some contexts, a robust prediction is needed: if the preference of an object over another is too small, a small perturbation could swap the order. Abstention of preference is therefore useful.
- A robust method to estimate τ^* already exists in the literature (Cheng, 2012). Our motivation is to add it the imprecise probabilities framework.

Precise case

- Local learning: prediction based on the nearest neighbours (Cheng, 2010). The probability to observe a ranking is then:

$$P(\tau_1, \ldots, \tau_K | v) = \prod_{i=1}^{K} \prod_{m=1}^{M_i} \frac{v_{\tau_i(m)}}{\sum_{j=m}^{M_i} v_{\tau_i(j)}}. \quad (1)$$

  v^* is found with the maximum likelihood estimation (MLE).

- Obtaining the maximum can be done with different optimisation algorithms. One of them is the minorization-maximization algorithm (Hunter, 2004), a generalisation of the EM algorithm, that provably converges to the maximum.
- τ^* can be found be simply ordering the labels according to v^*.

Imprecise case

- An IPL model is in correspondence one-to-one with an imprecise parameter estimate, found by extending of the classical likelihood with the contour likelihood function (Edwards, 1992).

$$L^*(v) = \frac{L(v)}{\max_{v \in \Sigma} L(v)};$$

  L^* takes values in $]0, 1]$: the closer $L^*(v)$ is to 1, the more likely v is.

- "Cuts" can be done to find imprecise estimates. Given β in $[0, 1]$, the β-cut of the contour likelihood, written B^*_β, is defined by

$$B^*_\beta = \{v \in \Sigma : L^*(v) \geq \beta\}.$$

Example of contour likelihood usage

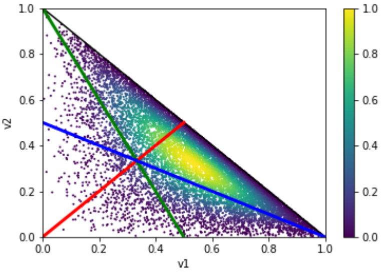


Figure: Contour likelihood

- Impossible to model the whole function: generation of strengths v with Dirichlet distributions.
- Threshold of 0.9 (yellow): all points indicate $\lambda_1 \succ \lambda_2 \succ \lambda_3$. This is the predicted ordering.
- Threshold of 0.5 (light-blue): some points indicate that $\lambda_2 \succ \lambda_1$ and $\lambda_3 \succ \lambda_2$. Only $\lambda_1 \succ \lambda_3$ is predicted.

Experimental method

- $CR(\pi, \hat{\pi}) = \frac{C-D}{C+D}$ the correctness and $CP(\pi, \hat{\pi}) = \frac{C+D}{n(n-1)/2}$ the completeness of predictions (Cheng, 2010). C concording pairs, D discording pairs. Metrics between 0 and 1.
- Tested on different data sets (Cheng, 2010). Examples: Bodyfat with $m = 252$ instances and $n = 7$ labels, Housing with $m = 506$, $n = 6$ and Wisconsin with $m = 194$, $n = 16$.
- Number of neighbours by cross validation. 200 strengths generated by two Dirichlet distributions: $\alpha_1 = \nu, \alpha_2 = 10\nu$. Cross validation for the results: 10 folds and repeated 3 times.
- Tested with different perturbations: missing elements in the rankings and swapped labels.

General results

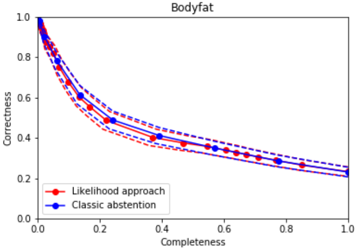


Figure: Full comparison of methods on Bodyfat with no missing labels

- Both methods seem to achieve similar results.
- Specific cases have to be identified: possible evolution when the amount of data for learning varies.

Results with missing data

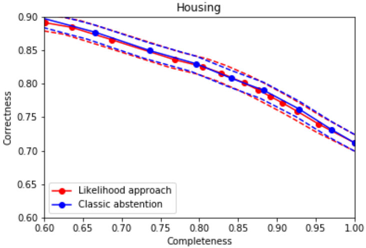


Figure: Comparison of methods on Housing with no missing labels

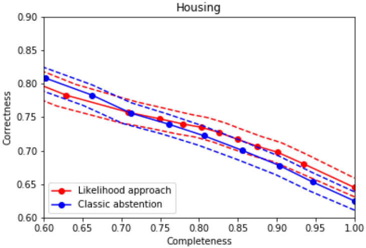


Figure: Comparison of methods on Housing with 60% of missing labels

Results with swapped labels

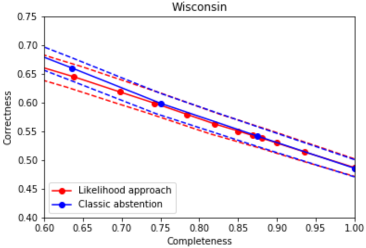


Figure: Comparison of methods on Wisconsin with no swapped labels

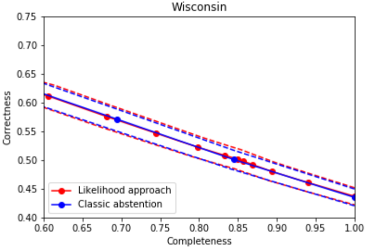


Figure: Comparison of methods on Wisconsin with 60% of swapped labels

Conclusions

We have introduced the Imprecise Plackett–Luce model (IPL).

We established algorithms to find (an outer approximation) of the maximal rankings and the E-admissible rankings of an IPL.

We have studied methods to learning an IPL and performed experiments about our techniques.

Thank you for your attention!

Loïc Adam, Arthur Van Camp, Sébastien Destercke & Benjamin Quost