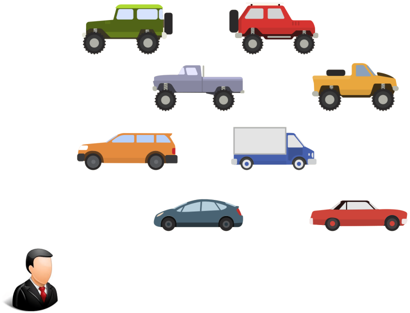
Preselection Bandits (Paper ID: 4941)

Viktor Bengs and Eyke Hüllermeier

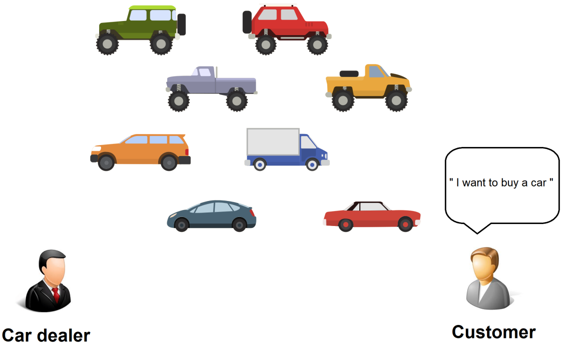
Heinz Nixdorf Institute
Paderborn University, Germany

Setting: Illustration



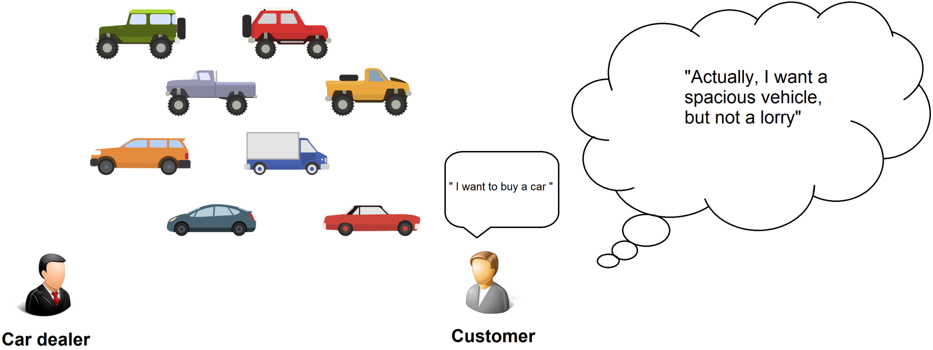
Car dealer

Setting: Illustration

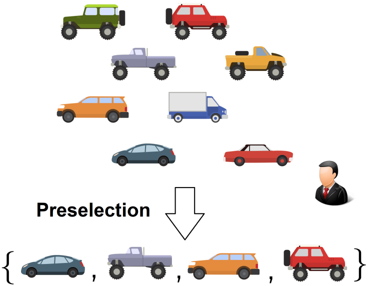


HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

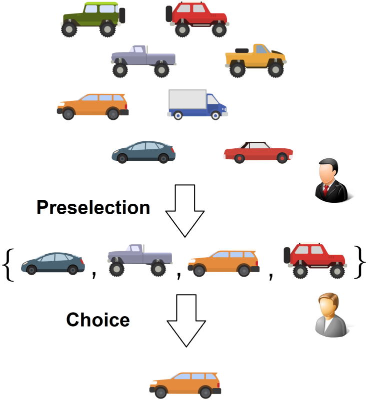
Setting: Illustration



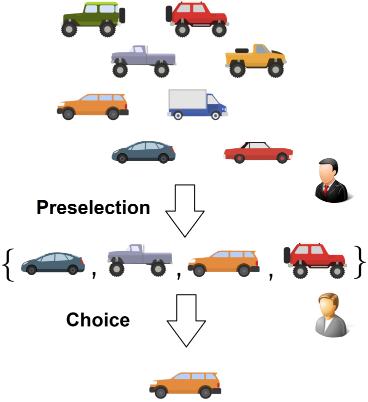
Setting: Illustration



Setting: Illustration



Setting: Illustration



→ How to make the preselection?

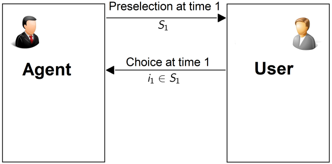
→ Which preselections lead to highly preferred choices?

Setting: Formalization

- Given n arms (choice options) $a_1, \ldots, a_n$
- An agent (learner) and a user (environment)
- A time period with time steps $1, 2, \ldots, T$

Setting: Formalization

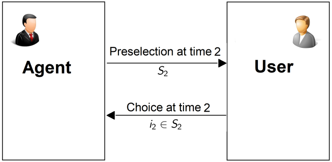
- Given n arms (choice options) $a_1, \ldots, a_n$
- An agent (learner) and a user (environment)
- A time period with time steps $1, 2, \ldots, T$
- In each time step $t \in \{1, 2, \ldots, T\}$:



  - ⋆ Agent preselects a subset $S_t \subset \{a_1, \ldots, a_n\}$ of arms for the user
  - ⋆ User chooses one arm from this subset S_t
  - ⋆ Chosen arm is replenished

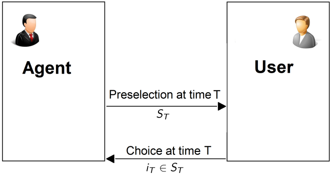
Setting: Formalization

- Given n arms (choice options) $a_1, \ldots, a_n$
- An agent (learner) and a user (environment)
- A time period with time steps $1, 2, \ldots, T$
- In each time step $t \in \{1, 2, \ldots, T\}$:
  - ⋆ Agent preselects a subset $S_t \subset \{a_1, \ldots, a_n\}$ of arms for the user
  - ⋆ User chooses one arm from this subset S_t
  - ⋆ Chosen arm is replenished



Setting: Formalization

- Given n arms (choice options) $a_1, \ldots, a_n$
- An agent (learner) and a user (environment)
- A time period with time steps $1, 2, \ldots, T$
- In each time step $t \in \{1, 2, \ldots, T\}$:

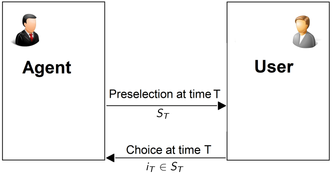


  - ⋆ Agent preselects a subset $S_t \subset \{a_1, \ldots, a_n\}$ of arms for the user
  - ⋆ User chooses one arm from this subset S_t
  - ⋆ Chosen arm is replenished

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Setting: Formalization

- Given n arms (choice options) $a_1, \ldots, a_n$
- An agent (learner) and a user (environment)
- A time period with time steps $1, 2, \ldots, T$
- In each time step $t \in \{1, 2, \ldots, T\}$:

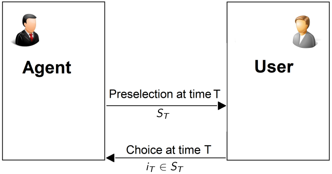


  - ⋆ Agent preselects a subset $S_t \subset \{a_1, \ldots, a_n\}$ of arms for the user
  - ⋆ User chooses one arm from this subset S_t
  - ⋆ Chosen arm is replenished

$\rightsquigarrow$ How exactly?

Setting: Formalization

- Given n arms (choice options) $a_1, \ldots, a_n$
- An agent (learner) and a user (environment)
- A time period with time steps $1, 2, \ldots, T$
- In each time step $t \in \{1, 2, \ldots, T\}$:



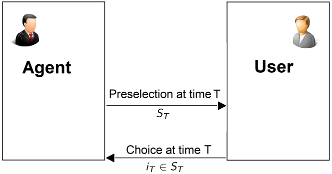
  - ⋆ Agent preselects a subset $S_t \subset \{a_1, \ldots, a_n\}$ of arms for the user
  - ⋆ User chooses one arm from this subset S_t
  - ⋆ Chosen arm is replenished

⇝ How exactly?

- ⋆ Assumption: User does **not** act perfectly w.r.t. its preferences
  $\Rightarrow$ Makes mistakes, i.e., chooses **not** mandatory the most preferred arm in a subset

Setting: Formalization

- Given n arms (choice options) $a_1, \dots, a_n$
- An agent (learner) and a user (environment)
- A time period with time steps $1, 2, \dots, T$
- In each time step $t \in \{1, 2, \dots, T\}$:

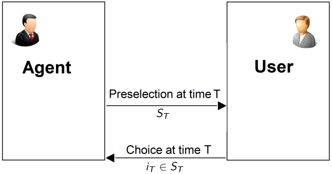


  ⋆ Agent preselects a subset $S_t \subset \{a_1, \dots, a_n\}$ of arms for the user
  ⋆ User chooses one arm from this subset S_t
  ⋆ Chosen arm is replenished

$\rightsquigarrow$ How exactly?

  ⋆ Assumption: User does **not** act perfectly w.r.t. its preferences
  $\Rightarrow$ Makes mistakes, i.e., chooses **not** mandatory the most preferred arm in a subset
  ⋆ Reasons: time constraints, incomplete information about the arms, ...

Setting: Formalization



- Given n arms (choice options) $a_1, \ldots, a_n$
- An agent (learner) and a user (environment)
- A time period with time steps $1, 2, \ldots, T$
- In each time step $t \in \{1, 2, \ldots, T\}$:
  - ⋆ Agent preselects a subset $S_t \subset \{a_1, \ldots, a_n\}$ of arms for the user
  - ⋆ User chooses one arm from this subset S_t
  - ⋆ Chosen arm is replenished

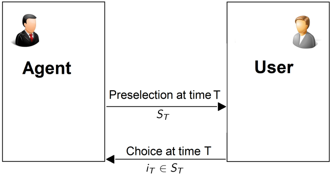
⇝ How exactly?

- ⋆ Assumption: User does **not** act perfectly w.r.t. its preferences
  ⇒ Makes mistakes, i.e., chooses **not** mandatory the most preferred arm in a subset
- ⋆ Reasons: time constraints, incomplete information about the arms, . . .

⇒ We model this choice behavior as an i.i.d. **random** process

Setting: Formalization

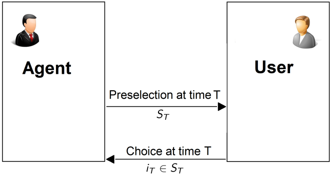
- Given n arms (choice options) $a_1, \ldots, a_n$
- An agent (learner) and a user (environment)
- A time period with time steps $1, 2, \ldots, T$
- In each time step $t \in \{1, 2, \ldots, T\}$:



  - ⋆ Agent preselects a subset $S_t \subset \{a_1, \ldots, a_n\}$ of arms for the user
  - ⋆ User chooses one arm from this subset S_t w.r.t. some random process
  - ⋆ Chosen arm is replenished

Setting: Formalization

- Given n arms (choice options) $a_1, \dots, a_n$
- An agent (learner) and a user (environment)
- A time period with time steps $1, 2, \dots, T$
- In each time step $t \in \{1, 2, \dots, T\}$:



  - ⋆ Agent preselects a subset $S_t \subset \{a_1, \dots, a_n\}$ of arms for the user
  - ⋆ User chooses one arm from this subset S_t w.r.t. some random process
  - ⋆ Chosen arm is replenished

Questions:

1. What is a good preselection of arms?
2. How can we learn it in an online learning framework?

Contribution

1st contribution: Introduction of the Preselection Bandits

Preselection Bandits: Exploration vs. exploitation of the optimal preselection

Contribution

1st contribution: Introduction of the Preselection Bandits

Preselection Bandits: Exploration vs. exploitation of the optimal preselection

⋆ Using preselections that have turned out to be appealing in the past

Contribution

1st contribution: Introduction of the Preselection Bandits

Preselection Bandits: Exploration vs. exploitation of the optimal preselection

- ⋆ Using preselections that have turned out to be appealing in the past
- ⋆ Trying out other, possibly even better preselections, but whose appeal is not precisely known so far

Contribution

1st contribution: Introduction of the Preselection Bandits

Preselection Bandits: Exploration vs. exploitation of the optimal preselection

- ⋆ Using preselections that have turned out to be appealing in the past
- ⋆ Trying out other, possibly even better preselections, but whose appeal is not precisely known so far

2nd contribution: Algorithmic solutions for two variants of the problem

Contribution

1st contribution: Introduction of the Preselection Bandits

Preselection Bandits: Exploration vs. exploitation of the optimal preselection

- ⋆ Using preselections that have turned out to be appealing in the past
- ⋆ Trying out other, possibly even better preselections, but whose appeal is not precisely known so far

2nd contribution: Algorithmic solutions for two variants of the problem

- Restricted Preselection – All subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ are admissible preselections

Contribution

1st contribution: Introduction of the Preselection Bandits

Preselection Bandits: Exploration vs. exploitation of the optimal preselection

- ⋆ Using preselections that have turned out to be appealing in the past
- ⋆ Trying out other, possibly even better preselections, but whose appeal is not precisely known so far

2nd contribution: Algorithmic solutions for two variants of the problem

- Restricted Preselection – All subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ are admissible preselections

$\rightsquigarrow$ Combinatorial challenging

Contribution

1st contribution: Introduction of the Preselection Bandits

Preselection Bandits: Exploration vs. exploitation of the optimal preselection

- ⋆ Using preselections that have turned out to be appealing in the past
- ⋆ Trying out other, possibly even better preselections, but whose appeal is not precisely known so far

2nd contribution: Algorithmic solutions for two variants of the problem

- Restricted Preselection – All subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ are admissible preselections

$\rightsquigarrow$ Combinatorial challenging

- Flexible Preselection – All non-empty subsets of arms are admissible preselections

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Contribution

1st contribution: Introduction of the Preselection Bandits

Preselection Bandits: Exploration vs. exploitation of the optimal preselection

- ⋆ Using preselections that have turned out to be appealing in the past
- ⋆ Trying out other, possibly even better preselections, but whose appeal is not precisely known so far

2nd contribution: Algorithmic solutions for two variants of the problem

- Restricted Preselection – All subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ are admissible preselections

$\rightsquigarrow$ Combinatorial challenging

- Flexible Preselection – All non-empty subsets of arms are admissible preselections

$\rightsquigarrow$ Trade-off: Size vs. information gain

Contribution

1st contribution: Introduction of the Preselection Bandits

Preselection Bandits: Exploration vs. exploitation of the optimal preselection

- ⋆ Using preselections that have turned out to be appealing in the past
- ⋆ Trying out other, possibly even better preselections, but whose appeal is not precisely known so far

2nd contribution: Algorithmic solutions for two variants of the problem

- Restricted Preselection – All subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ are admissible preselections

$\rightsquigarrow$ Combinatorial challenging

- Flexible Preselection – All non-empty subsets of arms are admissible preselections

$\rightsquigarrow$ Trade-off: Size vs. information gain

Related settings: Battling Bandits (Saha and Gopalan, 2018, 2019), Stochastic click models (Zoghi et al., 2017; Lattimore et al., 2018), MNL Bandits (Agrawal et al., 2016, 2017)

User Choice Modeling

Let us specify: "User chooses one arm from S w.r.t. some random process"

User Choice Modeling

Let us specify: "User chooses one arm from S w.r.t. some random process"

arms $\quad a_1 \quad a_2 \quad \ldots \quad a_n$

User Choice Modeling

Let us specify: "User chooses one arm from S w.r.t. some random process"

$$\begin{array}{lcccc} \text{arms} & a_1 & a_2 & \dots & a_n \\ & \Downarrow & \Downarrow & & \Downarrow \\ \text{strengths} & \theta_1 & \theta_2 & \dots & \theta_n \end{array}$$

User Choice Modeling

Let us specify: "User chooses one arm from S w.r.t. some random process"

$$
\begin{array}{ccccc}
\text{arms} & a_1 & a_2 & \dots & a_n \\
 & \Downarrow & \Downarrow & & \Downarrow \\
\text{strengths} & \theta_1 & \theta_2 & \dots & \theta_n
\end{array}
$$

- $\theta_i \in \mathbb{R}_+$ is the strength of arm a_i

User Choice Modeling

Let us specify: "User chooses one arm from S w.r.t. some random process"

$$\begin{array}{lcccc} \text{arms} & a_1 & a_2 & \dots & a_n \\ & \Downarrow & \Downarrow & & \Downarrow \\ \text{strengths} & \theta_1 & \theta_2 & \dots & \theta_n \end{array}$$

- $\theta_i \in \mathbb{R}_+$ is the strength of arm a_i
- Pr(user chooses arm $a_i \in S$) $= \frac{\theta_i}{\sum_{a_j \in S} \theta_j}$

  $\Rightarrow$ Probability for choosing a_i is proportional to its strength (MNL model)

User Choice Modeling: Preciseness

- Suppose θ_i is decomposable as $\theta_i = v_i^{\gamma}$
  - ⋆ $v_i \in \mathbb{R}_+$: latent utility of arm a_i (w.l.o.g. $v_i \neq v_j$ if $i \neq j$)
  - ⋆ $\gamma \in (0, \infty)$: degree of user's preciseness

User Choice Modeling: Preciseness

- Suppose θ_i is decomposable as $\theta_i = v_i^\gamma$
  - ⋆ $v_i \in \mathbb{R}_+$: latent utility of arm a_i (w.l.o.g. $v_i \neq v_j$ if $i \neq j$)
  - ⋆ $\gamma \in (0, \infty)$: degree of user's preciseness

$$\Pr(\text{user chooses arm } a_i \in S) = \frac{\theta_i}{\sum_{a_j \in S} \theta_j}$$

User Choice Modeling: Preciseness

- Suppose θ_i is decomposable as $\theta_i = v_i^\gamma$
  - ⋆ $v_i \in \mathbb{R}_+$: latent utility of arm a_i (w.l.o.g. $v_i \neq v_j$ if $i \neq j$)
  - ⋆ $\gamma \in (0, \infty)$: degree of user's preciseness

$$\Pr(\text{user chooses arm } a_i \in S) = \frac{\theta_i}{\sum_{a_j \in S} \theta_j} = \frac{v_i^\gamma}{\sum_{a_j \in S} v_j^\gamma}$$

User Choice Modeling: Preciseness

- Suppose θ_i is decomposable as $\theta_i = v_i^\gamma$
    - ⋆ $v_i \in \mathbb{R}_+$: latent utility of arm a_i (w.l.o.g. $v_i \neq v_j$ if $i \neq j$)
    - ⋆ $\gamma \in (0, \infty)$: degree of user's preciseness

$$\Pr(\text{user chooses arm } a_i \in S) = \frac{\theta_i}{\sum_{a_j \in S} \theta_j} = \frac{v_i^\gamma}{\sum_{a_j \in S} v_j^\gamma} \to \begin{cases} \frac{1}{|S|}, & \lim\limits_{\gamma \searrow 0}, \\ & \end{cases}$$

User Choice Modeling: Preciseness

- Suppose θ_i is decomposable as $\theta_i = v_i^\gamma$
  - ⋆ $v_i \in \mathbb{R}_+$: latent utility of arm a_i (w.l.o.g. $v_i \neq v_j$ if $i \neq j$)
  - ⋆ $\gamma \in (0, \infty)$: degree of user's preciseness

$$\Pr(\text{user chooses arm } a_i \in S) = \frac{\theta_i}{\sum_{a_j \in S} \theta_j} = \frac{v_i^\gamma}{\sum_{a_j \in S} v_j^\gamma} \to \begin{cases} \frac{1}{|S|}, & \lim\limits_{\gamma \searrow 0}, \\ \mathbb{1}_{a_i = \operatorname{argmax}_{a_j \in S} v_j}, & \lim\limits_{\gamma \nearrow \infty} \end{cases}.$$

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

User Choice Modeling: Preciseness

- Suppose θ_i is decomposable as $\theta_i = v_i^\gamma$
  - ⋆ $v_i \in \mathbb{R}_+$: latent utility of arm a_i (w.l.o.g. $v_i \neq v_j$ if $i \neq j$)
  - ⋆ $\gamma \in (0, \infty)$: degree of user's preciseness

$$\Pr(\text{user chooses arm } a_i \in S) = \frac{\theta_i}{\sum_{a_j \in S} \theta_j} = \frac{v_i^\gamma}{\sum_{a_j \in S} v_j^\gamma} \to \begin{cases} \frac{1}{|S|}, & \lim_{\gamma \searrow 0}, \\ \mathbb{1}_{a_i = \underset{a_j \in S}{\mathrm{argmax}}\ v_j}, & \lim_{\gamma \nearrow \infty} \end{cases}.$$

$$\Rightarrow \begin{array}{ll} \gamma \text{ small:} & \text{imprecise user} \\ \gamma \text{ large:} & \text{precise user} \end{array}$$

Optimal Preselection

1st Question: What is a good preselection of arms?

- Suppose latent utilities $v_1, v_2, \ldots, v_n$ were known

Optimal Preselection

1st Question: What is a good preselection of arms?

- Suppose latent utilities $v_1, v_2, \ldots, v_n$ were known

Optimal Preselection

1st Question: What is a good preselection of arms?

- Suppose latent utilities $v_1, v_2, \ldots, v_n$ were known
- Compute expected utility of a set $S \subset \{a_1, \ldots, a_n\}$:

$$\mathcal{U}(S) = \sum_{a_i \in S} v_i \cdot \Pr(\text{user chooses arm } a_i \in S)$$

Optimal Preselection

1st Question: What is a good preselection of arms?

- Suppose latent utilities $v_1, v_2, \ldots, v_n$ were known
- Compute expected utility of a set $S \subset \{a_1, \ldots, a_n\}$:

$$\mathcal{U}(S) = \sum_{a_i \in S} v_i \cdot \Pr(\text{user chooses arm } a_i \in S) = \sum_{a_i \in S} v_i \cdot \frac{v_i^\gamma}{\sum_{a_j \in S} v_j^\gamma} = \frac{\sum_{a_i \in S} v_i^{1+\gamma}}{\sum_{a_j \in S} v_j^\gamma}$$

Optimal Preselection

1st Question: What is a good preselection of arms?

- Suppose latent utilities $v_1, v_2, \dots, v_n$ were known
- Compute expected utility of a set $S \subset \{a_1, \dots, a_n\}$:

$$\mathcal{U}(S) = \sum_{a_i \in S} v_i \cdot \Pr(\text{user chooses arm } a_i \in S) = \sum_{a_i \in S} v_i \cdot \frac{v_i^{\gamma}}{\sum_{a_j \in S} v_j^{\gamma}} = \frac{\sum_{a_i \in S} v_i^{1+\gamma}}{\sum_{a_j \in S} v_j^{\gamma}}$$

- $\mathcal{S}$: Set of admissible preselections

$\Rightarrow$ Optimal preselection is $S^* = \underset{S \in \mathcal{S}}{argmax}\, \mathcal{U}(S)$

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Optimal Preselection

1st Question: What is a good preselection of arms?

- Suppose latent utilities $v_1, v_2, \ldots, v_n$ were known
- Compute expected utility of a set $S \subset \{a_1, \ldots, a_n\}$:

$$\mathcal{U}(S) = \sum_{a_i \in S} v_i \cdot \Pr(\text{user chooses arm } a_i \in S) = \sum_{a_i \in S} v_i \cdot \frac{v_i^{\gamma}}{\sum_{a_j \in S} v_j^{\gamma}} = \frac{\sum_{a_i \in S} v_i^{1+\gamma}}{\sum_{a_j \in S} v_j^{\gamma}}$$

- $\mathcal{S}$: Set of admissible preselections

$\Rightarrow$ Optimal preselection is $S^* = \underset{S \in \mathcal{S}}{argmax}\, \mathcal{U}(S)$

$\Rightarrow$ Performance measure for learner: Regret at time t, i.e., $\mathcal{U}(S^*) - \mathcal{U}(S_t)$

Variants of Preselection

1st Question: What is a good preselection of arms?

It depends on the set of admissible preselections $\mathcal{S}$:

Variants of Preselection

1st Question: What is a good preselection of arms?

It depends on the set of admissible preselections $\mathcal{S}$:

- If $\mathcal{S}$ contains all non-empty subsets of arms (Flexible Preselection)

Variants of Preselection

1st Question: What is a good preselection of arms?

It depends on the set of admissible preselections $\mathcal{S}$:

- If $\mathcal{S}$ contains all non-empty subsets of arms (Flexible Preselection)

$\rightsquigarrow$ The optimal preselection S^* is just the single best arm

Variants of Preselection

1st Question: What is a good preselection of arms?

It depends on the set of admissible preselections $\mathcal{S}$:

- If $\mathcal{S}$ contains all non-empty subsets of arms (Flexible Preselection)

$\rightsquigarrow$ The optimal preselection S^* is just the single best arm

- If $\mathcal{S}$ contains all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ (Restricted Preselection)

Variants of Preselection

1st Question: What is a good preselection of arms?

It depends on the set of admissible preselections $\mathcal{S}$:

- If $\mathcal{S}$ contains all non-empty subsets of arms (Flexible Preselection)

$\rightsquigarrow$ The optimal preselection S^* is just the single best arm

- If $\mathcal{S}$ contains all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ (Restricted Preselection)
    - $\star$ Due to γ not necessarily consisting of the best arms!

Variants of Preselection

1st Question: What is a good preselection of arms?

It depends on the set of admissible preselections $\mathcal{S}$:

- If $\mathcal{S}$ contains all non-empty subsets of arms (Flexible Preselection)

$\rightsquigarrow$ The optimal preselection S^* is just the single best arm

- If $\mathcal{S}$ contains all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ (Restricted Preselection)
  - $\star$ Due to γ not necessarily consisting of the best arms!

$\rightsquigarrow$ small γ (imprecise user) $\rightsquigarrow$ S^* has to consist of the best arms

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Variants of Preselection

1st Question: What is a good preselection of arms?

It depends on the set of admissible preselections $\mathcal{S}$:

- If $\mathcal{S}$ contains all non-empty subsets of arms (Flexible Preselection)

$\rightsquigarrow$ The optimal preselection S^* is just the single best arm

- If $\mathcal{S}$ contains all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ (Restricted Preselection)
  - $\star$ Due to γ not necessarily consisting of the best arms!

$$\rightsquigarrow \quad \begin{array}{cccc} \text{small } \gamma & \text{(imprecise user)} & \rightsquigarrow & S^* \text{ has to consist of the best arms} \\ \text{large } \gamma & \text{(precise user)} & \rightsquigarrow & S^* \text{ has only to entail the best arm} \end{array}$$

Variants of Preselection

1st Question: What is a good preselection of arms?

It depends on the set of admissible preselections $\mathcal{S}$:

- If $\mathcal{S}$ contains all non-empty subsets of arms (Flexible Preselection)

$\rightsquigarrow$ The optimal preselection S^* is just the single best arm

- If $\mathcal{S}$ contains all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ (Restricted Preselection)
  - $\star$ Due to γ not necessarily consisting of the best arms!

$$\rightsquigarrow \begin{array}{ccccc} \text{small } \gamma & \text{(imprecise user)} & \rightsquigarrow & S^* \text{ has to consist of the best arms} \\ \text{large } \gamma & \text{(precise user)} & \rightsquigarrow & S^* \text{ has only to entail the best arm} \end{array}$$

  - $\star$ In general: Composition of best and worst arms

Variants of Preselection

1st Question: What is a good preselection of arms?

It depends on the set of admissible preselections $\mathcal{S}$:

- If $\mathcal{S}$ contains all non-empty subsets of arms (Flexible Preselection)

$\rightsquigarrow$ The optimal preselection S^* is just the single best arm

- If $\mathcal{S}$ contains all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ (Restricted Preselection)
  - $\star$ Due to γ not necessarily consisting of the best arms!

$$\rightsquigarrow \quad \begin{array}{ccccc} \text{small } \gamma & \text{(imprecise user)} & \rightsquigarrow & S^* \text{ has to consist of the best arms} \\ \text{large } \gamma & \text{(precise user)} & \rightsquigarrow & S^* \text{ has only to entail the best arm} \end{array}$$

  - $\star$ In general: Composition of best and worst arms

$\rightsquigarrow$ Allows to capture decision-making biases of users ("decoy effect")

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

In the following: Restricted preselections ($\mathcal{S}$ = all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$)

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

In the following: Restricted preselections ($\mathcal{S} =$ all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$)

- Problem: Difficult to learn latent utilities $v_1, v_2, \ldots, v_n$ directly

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

In the following: Restricted preselections ($\mathcal{S} =$ all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$)

- Problem: Difficult to learn latent utilities $v_1, v_2, \ldots, v_n$ directly
- *Use relative utilities instead:* a_J is some reference arm

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

In the following: Restricted preselections ($\mathcal{S}$ = all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$)

- Problem: Difficult to learn latent utilities $v_1, v_2, \ldots, v_n$ directly
- *Use relative utilities instead:* a_J is some reference arm

$$v_1, v_2, \ldots, v_n \quad \rightarrow \quad \underbrace{\left(v_1/v_J\right)^{\gamma}}_{O_{1,J}}, \underbrace{\left(v_2/v_J\right)^{\gamma}}_{O_{2,J}}, \ldots \underbrace{\left(v_n/v_J\right)^{\gamma}}_{O_{n,J}}$$

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

In the following: Restricted preselections ($\mathcal{S}$ = all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$)

- Problem: Difficult to learn latent utilities $v_1, v_2, \ldots, v_n$ directly
- *Use relative utilities instead:* a_J is some reference arm

$$v_1, v_2, \ldots, v_n \quad \rightarrow \quad \underbrace{\left(v_1/v_J\right)^\gamma}_{O_{1,J}}, \underbrace{\left(v_2/v_J\right)^\gamma}_{O_{2,J}}, \ldots \underbrace{\left(v_n/v_J\right)^\gamma}_{O_{n,J}} \quad \rightsquigarrow \text{Why?}$$

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

In the following: Restricted preselections ($\mathcal{S}$ = all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$)

- Problem: Difficult to learn latent utilities $v_1, v_2, \ldots, v_n$ directly
- *Use relative utilities instead:* a_J is some reference arm

$$v_1, v_2, \ldots, v_n \quad \rightarrow \quad \underbrace{\left({}^{v_1}/{}_{v_J}\right)^{\gamma}}_{O_{1,J}}, \underbrace{\left({}^{v_2}/{}_{v_J}\right)^{\gamma}}_{O_{2,J}}, \ldots \underbrace{\left({}^{v_n}/{}_{v_J}\right)^{\gamma}}_{O_{n,J}} \quad \rightsquigarrow \text{Why?}$$

$$\Rightarrow \quad \underset{S}{\mathit{argmax}}\, \mathcal{U}(S; (v_i)_{i=1}^n) \quad = \quad \underset{S}{\mathit{argmax}}\, \mathcal{U}(S; (O_{i,J}^{1/\gamma})_{i=1}^n)$$

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

In the following: Restricted preselections ($\mathcal{S}$ = all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$)

- Problem: Difficult to learn latent utilities $v_1, v_2, \ldots, v_n$ directly
- *Use relative utilities instead:* a_J is some reference arm

$$v_1, v_2, \ldots, v_n \quad \rightarrow \quad \underbrace{\left({}^{v_1}/_{v_J}\right)^{\gamma}}_{O_{1,J}}, \underbrace{\left({}^{v_2}/_{v_J}\right)^{\gamma}}_{O_{2,J}}, \ldots \underbrace{\left({}^{v_n}/_{v_J}\right)^{\gamma}}_{O_{n,J}} \quad \rightsquigarrow \text{Why?}$$

$$\Rightarrow \quad \operatorname*{argmax}_S \mathcal{U}(S; (v_i)_{i=1}^n) \quad = \quad \operatorname*{argmax}_S \mathcal{U}(S; (O_{i,J}^{1/\gamma})_{i=1}^n)$$

$\rightsquigarrow$ How to learn the $O_{i,J}$'s?

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

In the following: Restricted preselections ($\mathcal{S}$ = all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$)

- Problem: Difficult to learn latent utilities $v_1, v_2, \ldots, v_n$ directly
- *Use relative utilities instead:* a_J is some reference arm

$$v_1, v_2, \ldots, v_n \quad \rightarrow \quad \underbrace{\left(v_1/v_J\right)^\gamma}_{O_{1,J}}, \underbrace{\left(v_2/v_J\right)^\gamma}_{O_{2,J}}, \ldots \underbrace{\left(v_n/v_J\right)^\gamma}_{O_{n,J}} \quad \rightsquigarrow \text{Why?}$$

$$\Rightarrow \quad \operatorname*{argmax}_S \mathcal{U}(S; (v_i)_{i=1}^n) \quad = \quad \operatorname*{argmax}_S \mathcal{U}(S; (O_{i,J}^{1/\gamma})_{i=1}^n)$$

$\rightsquigarrow$ How to learn the $O_{i,J}$'s? $\overset{\text{MNL model}}{\Longrightarrow}$ $\quad O_{i,J} = \left(\frac{v_i}{v_J}\right)^\gamma \approx \frac{w_{i,J}}{w_{J,i}} =: \hat{O}_{i,J}$

$w_{i,j}$ = no. of times arm a_i was chosen, while a_j was also present in the preselection

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

In the following: Restricted preselections ($\mathcal{S}$ = all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$)

- Problem: Difficult to learn latent utilities $v_1, v_2, \ldots, v_n$ directly
- *Use relative utilities instead:* a_J is some reference arm

$$v_1, v_2, \ldots, v_n \quad \rightarrow \quad \underbrace{\left({}^{v_1}/{}_{v_J}\right)^{\gamma}}_{O_{1,J}}, \underbrace{\left({}^{v_2}/{}_{v_J}\right)^{\gamma}}_{O_{2,J}}, \ldots \underbrace{\left({}^{v_n}/{}_{v_J}\right)^{\gamma}}_{O_{n,J}} \quad \rightsquigarrow \text{Why?}$$

$$\Rightarrow \quad \underset{S}{\mathrm{argmax}}\, \mathcal{U}(S; (v_i)_{i=1}^n) \quad = \quad \underset{S}{\mathrm{argmax}}\, \mathcal{U}(S; (O_{i,J}^{1/\gamma})_{i=1}^n)$$

$\rightsquigarrow$ How to learn the $O_{i,J}$'s? $\overset{\text{MNL model}}{\Longrightarrow}$ $O_{i,J} = \left(\frac{v_i}{v_J}\right)^{\gamma} \approx \frac{w_{i,J}}{w_{J,i}} =: \hat{O}_{i,J}$

$w_{i,j}$ = no. of times arm a_i was chosen, while a_j was also present in the preselection

$\rightsquigarrow$ Which reference arm a_J?

Online Learner for Restricted Preselection

2nd Question: How to learn the optimal preselection?

In the following: Restricted preselections ($\mathcal{S}$ = all subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$)

- Problem: Difficult to learn latent utilities $v_1, v_2, \ldots, v_n$ directly
- *Use relative utilities instead:* a_J is some reference arm

$$v_1, v_2, \ldots, v_n \quad \rightarrow \quad \underbrace{\left(v_1/v_J\right)^{\gamma}}_{O_{1,J}}, \underbrace{\left(v_2/v_J\right)^{\gamma}}_{O_{2,J}}, \ldots \underbrace{\left(v_n/v_J\right)^{\gamma}}_{O_{n,J}} \quad \rightsquigarrow \text{Why?}$$

$$\Rightarrow \quad \underset{S}{argmax}\, \mathcal{U}(S; (v_i)_{i=1}^n) \quad = \quad \underset{S}{argmax}\, \mathcal{U}(S; (O_{i,J}^{1/\gamma})_{i=1}^n)$$

$\rightsquigarrow$ How to learn the $O_{i,J}$'s? $\overset{\text{MNL model}}{\Longrightarrow}$ $O_{i,J} = \left(\frac{v_i}{v_J}\right)^{\gamma} \approx \frac{w_{i,J}}{w_{J,i}} =: \hat{O}_{i,J}$

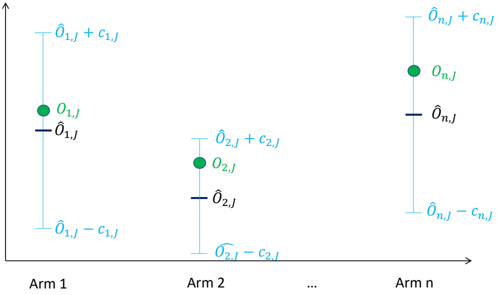
$w_{i,j}$ = no. of times arm a_i was chosen, while a_j was also present in the preselection

$\rightsquigarrow$ Which reference arm a_J? $\Longrightarrow$ The empirically most chosen so far!

The TRCB algorithm

Our algorithmic solution: Thresholding-Random-Confidence-Bound (TRCB) algorithm

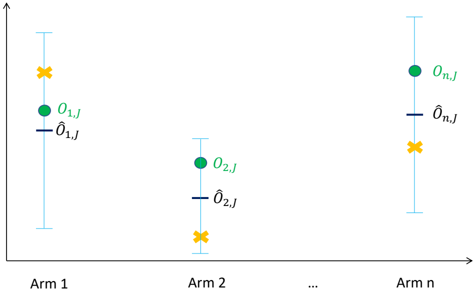
- For each relative utility $O_{i,J}$, compute confidence region based on $\hat{O}_{i,J}$



The TRCB algorithm

Our algorithmic solution: Thresholding-Random-Confidence-Bound (TRCB) algorithm

- Sample a random value inside each confidence region

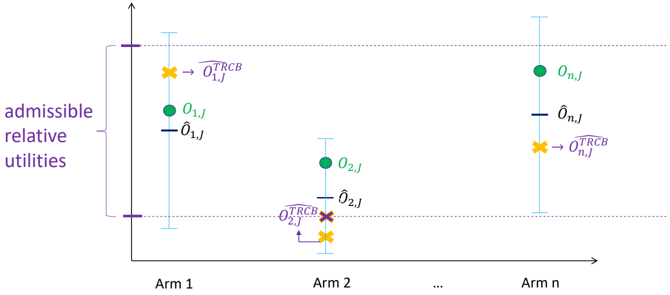


HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

The TRCB algorithm

Our algorithmic solution: Thresholding-Random-Confidence-Bound (TRCB) algorithm

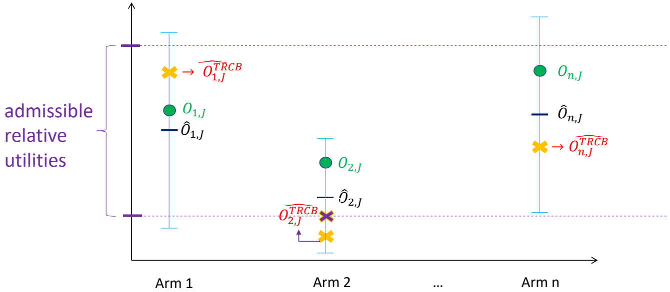
- Threshold the sampled values (to prevent inconsistencies)



The TRCB algorithm

Our algorithmic solution: Thresholding-Random-Confidence-Bound (TRCB) algorithm

- Threshold the sampled values (to prevent inconsistencies)

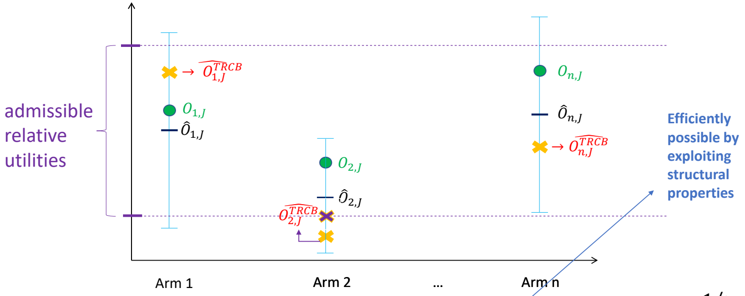


- maximize sampled expected utility $S_t = argmax_S\ \mathcal{U}(S; \left(\widehat{O_J^{TRCB}}\right)^{1/\gamma})$

The TRCB algorithm

Our algorithmic solution: Thresholding-Random-Confidence-Bound (TRCB) algorithm

- Threshold the sampled values (to prevent inconsistencies)



- maximize sampled expected utility $S_t = argmax_S\, \mathcal{U}(S; (\widehat{O_J^{TRCB}})^{1/\gamma})$

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Contribution (continued)

Our contribution: Introduction of the Preselection Bandits

- Restricted Preselection – All subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ are admissible preselections
- ⋆ Our suggestion: The Thresholding-Random-Confidence-Bound (TRCB) algorithm
- ⋆ Upper bound on cumulative regret: $O(\sqrt{n\,T}\,\log(T))$
- ⋆ Lower bound: $\Omega(\sqrt{n\,T})$

- Flexible Preselection – All non-empty subsets of arms are admissible preselections
- ⋆ Our suggestion (in the paper): The Confidence-Bound-Racing (CBR) algorithm
- ⋆ Upper bound on cumulative regret: $O(n\,\log(T))$
- ⋆ Lower bound: $\Omega(n\,\log(T))$

Contribution (continued)

Our contribution: Introduction of the Preselection Bandits

- Restricted Preselection – All subsets of a fixed size $k \in \mathbb{N}_{\geq 2}$ are admissible preselections

⋆ Our suggestion: The Thresholding-Random-Confidence-Bound (TRCB) algorithm

⋆ Upper bound on cumulative regret: $O(\sqrt{n\,T}\,\log(T))$

⋆ Lower bound: $\Omega(\sqrt{n\,T})$

- Flexible Preselection – All non-empty subsets of arms are admissible preselections

⋆ Our suggestion (in the paper): The Confidence-Bound-Racing (CBR) algorithm

⋆ Upper bound on cumulative regret: $O(n\,\log(T))$

⋆ Lower bound: $\Omega(n\,\log(T))$

+ Experimental study in the paper

Further research questions

- What if the strength parameters $\theta_1, \theta_2, \ldots, \theta_n$ also depend on the current user j?

Further research questions

- What if the strength parameters $\theta_1, \theta_2, \ldots, \theta_n$ also depend on the current user j?

- Can we incorporate feature information about the arm x_i and feature information about the user y_j?

Further research questions

- What if the strength parameters $\theta_1, \theta_2, \ldots, \theta_n$ also depend on the current user j?

- Can we incorporate feature information about the arm x_i and feature information about the user y_j?

- Different choice models

References

Agrawal, S., V. Avadhanula, V. Goyal, and A. Zeevi (2016). A near-optimal exploration-exploitation approach for assortment selection. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 599–600.

Agrawal, S., V. Avadhanula, V. Goyal, and A. Zeevi (2017). Thompson sampling for the mnl-bandit. In *Conference on Learning Theory*, pp. 76–78.

Lattimore, T., B. Kveton, S. Li, and C. Szepesvari (2018). Toprank: A practical algorithm for online stochastic ranking. In *Advances in Neural Information Processing Systems*, pp. 3945–3954.

Saha, A. and A. Gopalan (2018). Battle of bandits. In *Uncertainty in Artificial Intelligence*.

Saha, A. and A. Gopalan (2019). Regret minimisation in multinomial logit bandits. *arXiv preprint arXiv:1903.00543*.

Zoghi, M., T. Tunys, M. Ghavamzadeh, B. Kveton, C. Szepesvari, and Z. Wen (2017). Online learning to rank in stochastic click models. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 4199–4208.