Identifying Copeland Winners in Dueling Bandits with Indifferences

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Viktor Bengs[a,b], Björn Haddenhorst[c], Eyke Hüllermeier[a,b]

[a]Institute of Informatics, University of Munich, Germany
[b]Munich Center for Machine Learning, Germany
[c]Department of Computer Science, Paderborn University

viktor.bengs@lmu.de, eyke@lmu.de, willem.waegeman@UGent.be

TL;DR

Extension of Copeland winner identification in dueling bandits for indifference feedback with novel lower bounds and a worst-case nearly optimal learning algorithm

DUELING BANDITS WITH INDIFFERENCES

Setting

- **Given:** Different arms (options) $a_1, \dots, a_n \iff 1, \dots, n \iff \mathcal{A}$
- **Action at time t:** Choose a pair of arms $i_t \in \mathcal{A}$ and $j_t \in \mathcal{A} \setminus \{i_t\}$
- **Observation at time t:**

  either $i_t \succ j_t$, i.e., arm i_t is strictly preferred over arm j_t
  or $i_t \prec j_t$, i.e., arm j_t is strictly preferred over arm i_t
  or $i_t \cong j_t$, i.e., neither i_t is strictly preferred over j_t nor the opposite *(indifference between i_t and j_t)*
- **Stochastic feedback assumption:** Each possible explicit observations is determined by one of the following matrices $P^{\succ}, P^{\prec}, P^{\cong} \in [0,1]^{n \times n}$:

$$P^{\succ}_{i_t,j_t} = \mathbb{P}(i_t \succ j_t) \qquad P^{\prec}_{i_t,j_t} = \mathbb{P}(i_t \prec j_t) \qquad P^{\cong}_{i_t,j_t} = \mathbb{P}(i_t \cong j_t)$$

$\rightsquigarrow$ A problem instance is characterized by $\mathbf{P} = ((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}))_{i<j}$

Goal

(i) Finding a Copeland winner (COWI), i.e., an element of

$$\mathcal{C}(\mathbf{P}) = \{i \in \mathcal{A} \,|\, \mathrm{CP}(\mathbf{P}, i) = \max_j \mathrm{CP}(\mathbf{P}, j)\},$$

where

$$\mathrm{CP}(\mathbf{P}, i) = \sum_{j \neq i} 1_{[\![P^{\succ}_{i,j} > \max\{P^{\prec}_{i,j}, P^{\cong}_{i,j}\}]\!]} + \frac{1}{2} \sum_{j \neq i} 1_{[\![P^{\cong}_{i,j} > \max\{P^{\succ}_{i,j}, P^{\prec}_{i,j}\}]\!]},$$

is the Copeland score of arm $i \in \mathcal{A}$

(ii) Conducting as few as possible duels (low sample complexity)

Formal Goal: For a given error bound $\delta \in (0,1)$ design algorithm A which

• uses $\tau^{\mathrm{A}}(\mathbf{P})$ duels in total	such that	$\mathbb{E}[\tau^{\mathrm{A}}(\mathbf{P})]$ is small
• returns $\hat{i} \in \mathcal{A}$	such that	$\mathbb{P}(\hat{i} \notin \mathcal{C}(\mathbf{P})) \le \delta$

for any problem instance $\mathbf{P}$.

REFERENCES

[1] Róbert Busa-Fekete, Balázs Szörényi, Paul Weng, Weiwei Cheng, and Eyke Hüllermeier. Top-k selection based on adaptive sampling of noisy preferences. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1094–1102, 2013.

[2] Shubham Anand Jain, Rohan Shah, Sanit Gupta, Denil Mehta, Inderjeet J Nair, Jian Vora, Sushil Khyalia, Sourav Das, Vinay J Ribeiro, and Shivaram Kalyanakrishnan. PAC mode estimation using PPR martingale confidence sequences. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 5815–5852. PMLR, 2022.

[3] Tanguy Urvoy, Fabrice Clerot, Raphael Féraud, and Sami Naamane. Generic exploration and k-armed voting bandits. In *Proceedings of International Conference on Machine Learning (ICML)*, pages 91–99, 2013.

LEARNING ALGORITHM

POCOWISTA

Idea of POtential COpeland WInner STays Algorithm (POCOWISTA):

1. Duel arm i_t having highest potentially Copeland score with arm j_t having highest current Copeland score
2. Conduct duel via efficient PPR-1V1 routine [2] to find mode of $(P^{\succ}_{i_t,j_t}, P^{\cong}_{i_t,j_t}, P^{\prec}_{i_t,j_t})$

Algorithm POCOWISTA

1: **Input:** Set of arms $\mathcal{A}$, error prob. $\delta \in (0,1)$
2: **Initialization:** $e \leftarrow 1$ and for each $i \in \mathcal{A}$ set
  $D(i) \leftarrow \{i\}$ (set of already compared arms)
  $\widehat{CP}(i) \leftarrow 0$ (current Copeland score)
  $\overline{CP}(i) \leftarrow n-1$ (potential Copeland score)
3: **while** $\nexists i$ s.t. $\widehat{CP}(i) \geq \overline{CP}(j) \forall j \in \mathcal{A} \setminus \{i\}$ **do**
4:   $i_e = \mathrm{argmax}_{i \in \mathcal{A}} \overline{CP}(i)$
5:   $j_e = \mathrm{argmax}_{j \in \mathcal{A} \setminus D(i_e)} \widehat{CP}(j)$
6:   $k \leftarrow$ PPR-1V1$(i_e, j_e, \delta / \binom{n}{2})$
7:   Scores-Update(i_e, j_e, k)
8:   $e \leftarrow e+1$
9: **end while**
10: **return** $\mathrm{argmax}_{i \in \mathcal{A}} \widehat{CP}(i)$

Algorithm Scores-Update

1: **Input:** Arms i, j, ternary decision $k \in \{1,2,3\}$
2: **if** $k = 1$ **then**
3:   $\widehat{CP}(i) \leftarrow \widehat{CP}(i) + 1$
4: **else if** $k = 2$ **then**
5:   $\widehat{CP}(i) \leftarrow \widehat{CP}(i) + 1/2$, $\widehat{CP}(j) \leftarrow \widehat{CP}(j) + 1/2$
6: **else**
7:   $\widehat{CP}(j) \leftarrow \widehat{CP}(j) + 1$
8: **end if**
9: $D(i) \leftarrow D(i) \cup \{j\}$, $D(j) \leftarrow D(j) \cup \{i\}$
10: $\overline{CP}(i) \leftarrow n - |D(i)| + \widehat{CP}(i)$
11: $\overline{CP}(j) \leftarrow n - |D(j)| + \widehat{CP}(j)$

TRA-POCOWISTA

What if the problem instance $\mathbf{P}$ is transitive?

Definition. $\mathbf{P}$ is *transitive* if for each distinct $i, j, k \in \mathcal{A}$ holds:

1. Transitivity of strict preference.
   If $P^{\succ}_{i,j} > \max(P^{\prec}_{i,j}, P^{\cong}_{i,j})$ and $P^{\succ}_{j,k} > \max(P^{\prec}_{j,k}, P^{\cong}_{j,k})$, then $P^{\succ}_{i,k} > \max(P^{\prec}_{i,k}, P^{\cong}_{i,k})$.
2. IP-transitivity.
   If $P^{\cong}_{i,j} > \max(P^{\prec}_{i,j}, P^{\succ}_{i,j})$ and $P^{\succ}_{j,k} > \max(P^{\prec}_{j,k}, P^{\cong}_{j,k})$, then $P^{\succ}_{i,k} > \max(P^{\prec}_{i,k}, P^{\cong}_{i,k})$.
3. PI-transitivity.
   If $P^{\succ}_{i,j} > \max\left(P^{\prec}_{i,j}, P^{\cong}_{i,j}\right)$and $P^{\cong}_{j,k} > \max\left(P^{\prec}_{j,k}, P^{\succ}_{j,k}\right)$, then $P^{\succ}_{i,k} > \max\left(P^{\prec}_{i,k}, P^{\cong}_{i,k}\right)$.
4. Transitivity of indifference.
   If $P^{\cong}_{i,j} > \max(P^{\prec}_{i,j}, P^{\succ}_{i,j})$ and $P^{\cong}_{j,k} > \max(P^{\prec}_{j,k}, P^{\succ}_{j,k})$, then $P^{\cong}_{i,k} > \max(P^{\prec}_{i,k}, P^{\succ}_{i,k})$.

$\Rightarrow$ Updates can be made more efficient

Algorithm TRA-POCOWISTA

1: **Input:** Set of arms $\mathcal{A}$, error prob. $\delta \in (0,1)$
2: **Initialization:** $e \leftarrow 1$ and for each $i \in \mathcal{A}$
  $D(i) \leftarrow \{i\}$, $\widehat{CP}(i) \leftarrow 0$, $\overline{CP}(i) \leftarrow n-1$
  $W(i) \leftarrow \emptyset$ (set of defeated arms)
  $I(i) \leftarrow \emptyset$ (set of indifferent arms)
  $L(i) \leftarrow \emptyset$ (set of superior arms)
3: **while** $\nexists i$ s.t. $\widehat{CP}(i) \geq \overline{CP}(j) \forall j \in \mathcal{A} \setminus \{i\}$ **do**
4:   $i_e = \mathrm{argmax}_{i \in \mathcal{A}} \overline{CP}(i)$
5:   $j_e = \mathrm{argmax}_{j \in \mathcal{A} \setminus D(i_e)} \widehat{CP}(j)$
6:   $k \leftarrow$ PPR-1V1$(i_e, j_e, \delta/n)$
7:   Transitive-Score-Update(i_e, j_e, k)
8:   $e \leftarrow e+1$
9: **end while**
10: **return** $\mathrm{argmax}_{i \in \mathcal{A}} \widehat{CP}(i)$

Algorithm Transitive-Score-Update

1: **Input:** Arms i, j, $k \in \{1,2,3\}$
2: **if** $k = 1$ **then**
3:   $\widehat{CP}(i) \leftarrow \widehat{CP}(i) + |W(j) \cup I(j)| + 1$
4:   $W(i) \leftarrow W(i) \cup W(j) \cup I(j) \cup \{j\}$
5:   $D(i) \leftarrow D(i) \cup W(j) \cup I(j) \cup \{j\}$
6:   $L(j) \leftarrow L(j) \cup L(i) \cup I(i) \cup \{i\}$
7:   $D(j) \leftarrow D(j) \cup L(i) \cup I(i) \cup \{i\}$
8: **else if** $k = 2$ **then**
9:   $\widehat{CP}(i) \leftarrow \widehat{CP}(i) + |W(j)| + 1/2(1 + |I(j)|)$
10:   $\widehat{CP}(j) \leftarrow \widehat{CP}(j) + |W(i)| + 1/2(1 + |I(i)|)$
11:   $W(i) \leftarrow W(i) \cup W(j), W(j) \leftarrow W(i)$
12:   $L(i) \leftarrow L(i) \cup L(j), L(j) \leftarrow L(i)$
13:   $I(i) \leftarrow I(i) \cup I(j) \cup \{j\}, I(j) \leftarrow I(i) \cup I(j) \cup \{i\}$
14:   $D(i) \leftarrow D(i) \cup D(j)$, $D(j) \leftarrow D(i)$
15: **else**
16:   Same as for $k = 1$ with i and j reversed
17: **end if**
18: Same steps as line 10 and 11 in Score-Update

THEORETICAL RESULTS

Lower bounds

Informal Version: For $\mathbf{P}$ with $\min_{i<j} |P^{(1)}_{i,j} - P^{(2)}_{i,j}| > \Delta$ the lower bounds are

$$\Omega(n^2/\Delta^2 \ln 1/\delta),$$

where $P^{(1)}_{i,j}, P^{(2)}_{i,j}, P^{(3)}_{i,j}$ are the order statistics of $P^{\succ}_{i,j}$, $P^{\cong}_{i,j}$ and $P^{\prec}_{i,j}$.

Formal Version: If A correctly identifies the COWI with confidence $1-\delta$, then

$$\mathbb{E}[\tau^{\mathrm{A}}(\mathbf{P})] \geq \ln \frac{1}{2.4\delta} \sum_{j \in \mathcal{A} \setminus \{i^*\}} C_j \min_{k \in L(j) \cup I(j)} \frac{1}{D_{j,k}(\mathbf{P})},$$

where $\mathcal{C}(\mathbf{P}) = \{i^*\}$ and in the case with indifferences

$$D_{j,k}(\mathbf{P}) := \max\{\mathrm{KL}^{(1)}_{j,k}, \mathrm{KL}^{(2)}_{j,k}\}$$
$$\mathrm{KL}^{(1)}_{j,k} = \mathrm{KL}((P^{\succ}_{j,k}, P^{\cong}_{j,k}, P^{\prec}_{j,k}), (P^{\cong}_{j,k}, P^{\succ}_{j,k}, P^{\prec}_{j,k})),$$
$$\mathrm{KL}^{(2)}_{j,k} = \mathrm{KL}((P^{\succ}_{j,k}, P^{\cong}_{j,k}, P^{\prec}_{j,k}), (P^{\prec}_{j,k}, P^{\cong}_{j,k}, P^{\succ}_{j,k})),$$
$$C_j = \max_{(i,l) \in \Psi(j)} \frac{\binom{|I(j)|}{i}\binom{|L(j)|}{l}}{\binom{|I(j)|-1}{i-1}\binom{|L(j)|}{l} 1_{[\![i \geq 1]\!]} + \binom{|I(j)|}{i}\binom{|L(j)|-1}{l-1} 1_{[\![l \geq 1]\!]}},$$
$$\Psi(j) := \left\{(i,l) \in \{0, \dots, |I(j)|\} \times \{0, \dots, |L(j)|\} \,|\, i + 2l \geq 2d_j + 1\right\}$$

for any $\mathbf{P}$ with $\min_{j,k} \min\{P^{\succ}_{j,k}, P^{\cong}_{j,k}, P^{\prec}_{j,k}\} > 0$.

Upper bounds

Informal Version: Worst-case sample complexities have the order

POCOWISTA	TRA-POCOWISTA*	SAVAGE** [3]	PBR-CCSO** [1]
$\frac{n^2}{\Delta^2_{i,j}} \ln\left(\frac{n}{\sqrt{\delta}} \cdot \frac{1}{\Delta_{i,j}}\right)$	$\frac{n}{\Delta^2_{i,j}} \ln\left(\frac{n}{\sqrt{\delta}} \cdot \frac{1}{\Delta_{i,j}}\right)$	$\frac{n^2}{\Delta^2_{i,j}} \ln\left(\frac{n}{\delta} \cdot \frac{1}{\Delta_{i,j}}\right)$	$\frac{n^2}{\Delta^2_{i,j}} \ln\left(\frac{n^2}{\delta} \cdot \frac{1}{\Delta_{i,j}}\right)$

*if $\mathbf{P}$ is transitive
** if there are no indifferences

Formal Version: For any $\mathbf{P} = ((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}))_{i<j}$, such that there exists no pair $i, j \in \mathcal{A}$ with $i \neq j$ and $P^{\succ}_{i,j} = P^{\succ}_{j,i} = 1/3$, it holds
(i) for A := POCOWISTA that

$$\mathbb{P}\big(\hat{i}_{\mathrm{A}} \in \mathcal{C}(\mathbf{P}) \text{ and } \tau^{\mathrm{A}}(\mathbf{P}) \leq t(\mathbf{P}, \delta)\big) \geq 1 - \delta,$$

where $t(\mathbf{P}, \delta) \leq \sum_{i<j} t_0\big((P^{\succ}_{i,j}, P^{\cong}_{i,j}, P^{\prec}_{i,j}), \delta/\binom{n}{2}\big)$,

$$t_0\big((p_1, p_2, p_3), \delta\big) = \frac{c_1 p_{(1)}}{(p_{(1)} - p_{(2)})^2} \ln\left(\frac{\sqrt{2\,c_2} p_{(1)}}{\sqrt{\delta}(p_{(1)} - p_{(2)})}\right), \tag{1}$$

$p_{(1)} \geq p_{(2)} \geq p_{(3)}$ is the order statistic of p_1, p_2, p_3, $c_1 = 194.07$, and $c_2 = 79.86$.
(ii) for A := TRA-POCOWISTA if $\mathbf{P}$ transitive that

$$\mathbb{P}\big(\hat{i}_{\mathrm{A}} \in \mathcal{C}(\mathbf{P}) \text{ and } \tau^{\mathrm{A}}(\mathbf{P}) \leq \tilde{t}(\mathbf{P}, \delta)\big) \geq 1 - \delta,$$

where $\tilde{t}(\mathbf{P}, \delta) = \sum_{e=1}^{E} t_0\big((P^{\succ}_{i_e,j_e}, P^{\cong}_{i_e,j_e}, P^{\prec}_{i_e,j_e}), \delta/n\big)$, t_0 is as in (1) and $E \leq n$.