PADERBORN UNIVERSITY

PADERBORN UNIVERSITY

PAC MAXING IN DUELING BANDITS UNDER STOCHASTIC TRANSITIVITY MODELS

SEMINAR PRESENTATION

Shivam Sharma

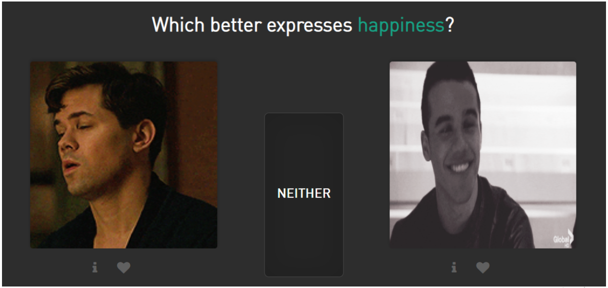
Paderborn, 18th Feb 2021

What is Maxing?

- Maxing or Maximum selection is a problem of identifying the best element from a set of different elements.
- This problem is a classical problem in computer science and has been revisited in machine learning.
- The identification of the best element is done through comparisons between the set elements.
- The best element is the one that wins all the comparisons against all the elements in the set.

Real life examples of Maxing

- A popular crowd-sourcing website called GIFGIF [`http://gifgif.media.mit.edu/`] is a perfect example for real life maxing example.
- The website uses maxing to assign emotions to the GIF image.



Dueling bandits

Note: In this presentation, we will use the terminology of a slot machine by considering elements as arms.

- Variataion of traditional Multi-armed bandit (MAB) problem.
- In MAB, pull an arm and observe the reward.
- In Dueling bandits setting, pull two different arms and only observe which arm gives the higher reward.

Need for PAC Maxing I

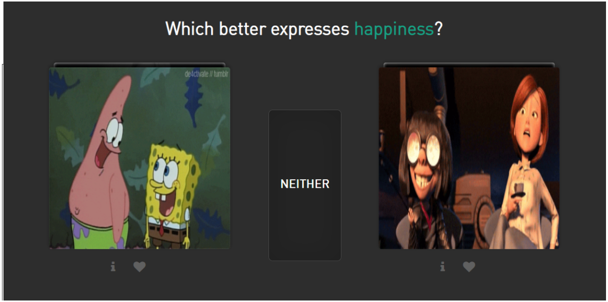


Figure: The best GIF image is the one which corresponds better to "Happiness"

Need for PAC Maxing II

- Difficult to find an exact winner from a set of arms.
- Instead, we can find an output that is a close approximation of the intended outcome.
- This setting is called Probably Approximately Correct (PAC) setting [1] and the maxing problem in this setting is called PAC maxing [2, 3, 4].

Notations I

To define the problem statement for PAC maxing, we will use the following notations.

- A fixed set of n arms $N = \{1, 2, \ldots, n\}$.
- Arms i and j are compared such that $1 \leq i \leq j \leq n$.
- The winner of the duel arm i (in this case) is returned with unknown pairwise probability $p_{i,j} \in [0, 1]$.
- $p_{j,i} = 1 - p_{i,j}$ (no ties) and $p_{i,i} = \frac{1}{2}$ holds true for all $i, j \in N$
- If $p_{i,j} > \frac{1}{2}$, i beats j. This defines an order $\succ$ over arms i and j.
- $\tilde{p}_{i,j} = p_{i,j} - \frac{1}{2}$ is the *calibrated pairwise probability* [5].
- $\tilde{p}_{i,j} = -\tilde{p}_{j,i}$.
- An arm i is called ϵ-preferable to j if $\tilde{p}_{i,j} \geq -\epsilon$.

Notations II

- If arm $i \in N$ is ϵ-preferable to all other arms, then it is called ϵ-maximum in N for $\epsilon \in (0, \frac{1}{2})$.
- An (ϵ, δ)-PAC maxing algorithm must output an ϵ-maximum arm with probability $\geq 1 - \delta$ for $\epsilon > 0$ and $0 < \delta \leq \frac{1}{2}$.
- ϵ is accuracy parameter and δ is confidence parameter.
- Arm i is called a *maximal* if $\tilde{p}_{i,j} > 0 \; \forall j \in N \setminus \{i\}$.

Why need restrictions?

Example 1

Probability preference matrix $P = [p_{i,j}]$, whose $(i,j)^{th}$ entry is $p_{i,j}$:

$$P = \begin{bmatrix} 0.5 & 0.7 & 0.1 \\ 0.3 & 0.5 & 0.6 \\ 0.9 & 0.4 & 0.5 \end{bmatrix}$$

Corresponding calibrated probability preference matrix [5] $\tilde{P} = [\tilde{p}_{i,j}]$, whose $(i,j)^{th}$ entry is the calibrated pairwise probability $\tilde{p}_{i,j} = p_{i,j} - \frac{1}{2}$:

$$\tilde{P} = \begin{bmatrix} 0 & 0.2 & -0.4 \\ -0.2 & 0 & 0.1 \\ 0.4 & -0.1 & 0 \end{bmatrix}$$

- No maximal in $\tilde{P}$.

Stochastic Transitivity I

The transitivities are given in papers [2, 3, 4].

- *Strong stochastic transitivity* (SST):
  $\forall$ distinct $i, j, k \in N : \tilde{p}_{i,j}, \tilde{p}_{j,k} \geq 0 \rightarrow \tilde{p}_{i,k} \geq \max\{\tilde{p}_{i,j}, \tilde{p}_{j,k}\}$
- *γ-relaxed stochastic transitivity* (γ-RST):
  For $\gamma \geq 1$ and $\forall$ distinct $i, j, k \in N : \tilde{p}_{i,j}, \tilde{p}_{j,k} \geq 0$
  $\rightarrow \gamma \cdot \tilde{p}_{i,k} \geq \max\{\tilde{p}_{i,j}, \tilde{p}_{j,k}\}$
- *Moderate stochastic transitivity* (MST):
  $\forall$ distinct $i, j, k \in N : \tilde{p}_{i,j}, \tilde{p}_{j,k} \geq 0 \rightarrow \tilde{p}_{i,k} \geq \min\{\tilde{p}_{i,j}, \tilde{p}_{j,k}\}$
- *Weak stochastic transitivity* (WST):
  $\forall$ distinct $i, j, k \in N : \tilde{p}_{i,j}, \tilde{p}_{j,k} \geq 0 \rightarrow \tilde{p}_{i,k} \geq 0$

Stochastic Transitivity II

Stochastic triangle inequality (STI):
$\forall$ distinct $i, j, k \in N : i \succ j \succ k \rightarrow \tilde{p}_{i,k} \leq \tilde{p}_{i,j} + \tilde{p}_{j,k}$

Example 2

$$\tilde{P} = \begin{bmatrix} 0 & 0.3 & 0.5 \\ -0.3 & 0 & 0.1 \\ -0.5 & -0.1 & 0 \end{bmatrix}$$

- SST is satisfied since $\tilde{p}_{1,2} \geq 0, \tilde{p}_{2,3} \geq 0$ and $\tilde{p}_{1,3} = 0.5 \geq \max\{0.1, 0.3\}$
- MST is also being satisfied since $\tilde{p}_{1,2} \geq 0, \tilde{p}_{2,3} \geq 0$ and $\tilde{p}_{1,3} \geq \min\{0.1, 0.3\}$.
- WST is also satisfied.

PAC Maxing I

Model	Maxing
SST, γ-RST with STI	SST: $\Theta(\frac{n}{\epsilon^2}\log\frac{1}{\delta})$ γ-RST: $\Theta(\frac{n\gamma^2}{\epsilon^2}\log\frac{1}{\delta})$
SST	$\Theta(\frac{n}{\epsilon^2}\log\frac{1}{\delta})$
MST with and without STI	$\Theta(\frac{n}{\epsilon^2}\log\frac{1}{\delta})^*$
WST with and without STI	$\Omega(n^2)$ $\mathcal{O}(\frac{n^2}{\epsilon^2}\log\frac{n}{\delta})$

Table: [4, Table 1]

*: for $\delta \geq \min(\frac{1}{n}, e^{-n^{1/4}})$

PAC Maxing for WST models with (+) and without (-) STI I

- Is PAC maxing for the most general models, i.e., the models with only WST property possible linearly with respect to the number of comparisons? [4]

Theorem (1)

There exists a model that satisfies WST for which any algorithm requires $\Omega(n^2)$ comparisons to find a $\frac{1}{4}$-maximum with probability $\geq \frac{7}{8}$ [4]

- There is **no** linearly complex PAC maxing algorithm possible for models with only WST property.
- BRUTE-FORCE [4] algorithm gives the upper bound on the sample complexity.
- It guarantees to estimate all pairwise probabilities to ϵ precision by using $\mathcal{O}(\frac{n^2}{\epsilon^2} \log \frac{n}{\delta})$ comparisons with probability $\geq 1 - \delta$ [4].

PAC Maxing for WST models with (+) and without (-) STI II

- The proof of it's correctness is based on Hoeffding's inequality and union bound [4].
- **Note:** The inclusion of STI property does not impact the quadratic complexity for maxing in WST models as seen in the Table 1 before [4].

PAC Maxing for MST models with and without STI I

- If a linear PAC maxing is not possible models under WST, is there any model, which is more general than SST and less general than WST, for which PAC maxing of linear complexity is possible? [4]
- OPT-MAX uses $\mathcal{O}(\frac{n}{\epsilon^2} \log \frac{1}{\delta})$ comparisons for maxing with probability $\geq 1 - \delta$, for $\delta \geq \min(\frac{1}{n}, e^{-n^{1/4}})$ [4].
- This makes MST the "most general" model **known** for which maxing is possible linearly [4, Section 1.1].

PAC Maxing for MST models with and without STI II

- OPT-MAX is a combination of three different algorithm for different ranges of confidence δ [4].
  - OPT-MAX-LOW for $\min(e^{-n^{1/4}}, \frac{1}{n}) \leq \delta \leq \frac{1}{n^{1/3}}$,
  - OPT-MAX-MEDIUM for $\frac{1}{n^{1/3}} \leq \delta \leq \frac{1}{\log n}$.
  - OPT-MAX-HIGH for $\frac{1}{\log n} \leq \delta$.
- The correctness of these three algorithms is based on union bound, which binds the number of comparisons used in all recursive calls in each algorithm [4].
- These three algorithms are finally merged into OPT-MAX.

Summary

- Models under MST assumption are the most general models **known** for which linearly complex PAC maxing algorithms are possible to find an ϵ-maximum arm with probability $\geq 1 - \delta$.
- The upper bound on the maxing under MST is the **strongest** bound among SST, γ-RST and MST.
- The correctness of most of the algorithms to find an ϵ-maximum arm from a set *N* with probability $\geq 1 - \delta$ is based on Hoeffding's inequality and union bound.

Q and A

Thank you for your attention.

References

[1] Eyal Even-Dar, Shie Mannor, and Yishay Mansour.
Pac bounds for multi-armed bandit and markov decision processes.
In *International Conference on Computational Learning Theory*, pages 255–270. Springer, 2002.

[2] Moein Falahatgar, Alon Orlitsky, Venkatadheeraj Pichapati, and Ananda Theertha Suresh.
Maximum selection and ranking under noisy comparisons.
arXiv preprint arXiv:1705.05366, 2017.

[3] Moein Falahatgar, Yi Hao, Alon Orlitsky, Venkatadheeraj Pichapati, and Vaishakh Ravindrakumar.
Maxing and ranking with few assumptions.
Advances in Neural Information Processing Systems, 30:7060–7070, 2017.

[4] Moein Falahatgar, Ayush Jain, Alon Orlitsky, Venkatadheeraj Pichapati, and Vaishakh Ravindrakumar.
The limits of maxing, ranking, and preference learning.
In *International Conference on Machine Learning*, pages 1427–1436, 2018.

[5] Viktor Bengs, Robert Busa-Fekete, Eyke Hüllermeier, and Adil El Mesaoudi-Paul.
Preference-based online learning with dueling bandits: A survey.
Under review.