Optimal Maxing and Sorting with Minimal Assumptions: It's Only a Game

With

Moein Falahatgar, Yi Hao, Ayush Jain, Dheeraj Pichapati
Vaishakh Ravi, Ananda Theertha Suresh

UCSD

NASIT, UCLA, August 2022

Optimal Maxing and Sorting with Minimal Assumptions: It's Only a Game

With

Moein Falahatgar, Yi Hao, Ayush Jain, Dheeraj Pichapati
Vaishakh Ravi, Ananda Theertha Suresh

UCSD

NASIT, UCLA, August 2022

The two most basic algorithms?

Given n elements

Use pairwise comparisons to:

- Find maximum $\longrightarrow$ Maxing
- Sort $\longrightarrow$ Ranking

\# comparisons needed

- N_{max}
- N_{sort}

Comparison models

- Deterministic
- Randomized

Close a few gaps

Pairwise comparisons always correct

$N_{\text{max}} = n - 1$

$\leq$ sequential elimination: pick candidate, compare, eliminate loser, ...

knockout: essentially same, on tree

$\geq$ graph, edges are comparisons performed

$n - 2$ comparisons $\longrightarrow$ graph disconnected $\longrightarrow$ ambiguity

$N_{\text{sort}} \sim n \log_2 n$

$\leq$ merge sort, quick sort, heap sort, $n \log_2 n$

$\geq \log_2 n! \sim \log_2 \left(\sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n\right) \sim n \log_2 n$, correct to constant factor

Comparisons answered randomly – error, noise

Suppose: i is larger than j

Ask: which larger

Deterministic

- i with probability 1
- j with probability 0

Random

- i with probability $p_{i,j}$
- j with probability $p_{j,i} = 1 - p_{i,j}$
- $p_{i,j}$ unknown, except $\geq \frac{1}{2}$

Repeated comparisons independent

$p_{i,j} = 0.6 \longrightarrow$ i, i, j, i, j

Increasing generality

Model	$p_{i,j}$
Fixed noise	$\frac{1}{2} \pm \epsilon$, all same ϵ
Parametric	Determined by n parameters, one for each element
Non-parametric	Essentially arbitrary

$0 \le \delta \le 1$ – allowed error

Want answer correct with probability $1 - \delta$

Max: Largest element w.p. $1 - \delta$

Sort: Correct order (all pairs) w.p. $1 - \delta$

	N_{max}	N_{sort}
Deterministic	$n - 1$	$\sim n \log_2 n$
Randomized	$\Theta(n)$?	$\Theta(n \log n)$?

Fixed Noise [Fei, Rag, Pel, Upf, '94]

Faulty circuits

$p_{i,j} = \frac{1}{2} \pm \epsilon$, same bias ϵ

Want: correct max, ranking with probability $\geq 1 - \delta$

$N_{\mathsf{max}} = \Theta\left(\frac{n}{\epsilon^2} \log \frac{1}{\delta}\right)$

For $\delta > \frac{1}{n}$, $\quad N_{\text{sort}} = \Theta\left(\frac{n \log n}{\epsilon^2}\right)$

Personal choice

Comparing easier than rating — drinks, vision, drivers, ICML papers

Advertising

Best ad to show on webpage — room for few (two?) ads, select best

Design

A/B testing

Information retrieval

Order of web-search links

Even classical... Sports

Best or sorted teams

NBA conference, league playoffs

Modern Times

Facts — alternative

Truth — relative

All opinion — where you come from

1990's — Circuits

$p_{i,j}$ — fixed value (0.9 or 0.1), known in advance, typically ≈ 1 or 0

2010's — Preferences

$p_{i,j}$ — arbitrary, not known in advance, can be $\approx \frac{1}{2}$

Sports — $P_{\text{Spain,Portugal}} \approx \frac{1}{2}$, $P_{\text{England,Panama}} \approx 1$

WSJ advertising — $P_{\text{BMW,Audi}} \approx \frac{1}{2}$, $P_{\text{BMW,Rickshaw}} \approx 1$

Need: Generalize fixed-noise model

Recall

- Elements $1, \ldots, n$
- $p_{i,j}$ - probability that i preferable to j
- $\binom{n}{2}$ probabilities

Parametric models

- Probabilities determined by just n parameters
- One associated with each of the n elements
- Parameters $x_1, \ldots, x_n \in \mathbb{R}$
- $f : \mathbb{R} \times \mathbb{R} \to [0, 1]$
- $p_{i,j} = f(x_i, x_j)$

n parameters viewed as "weights"

$w_i \geq 0$

$p_{i,j} = f(w_i, w_j) \overset{\text{def}}{=} \frac{w_i}{w_i + w_j}$

- $w_1 = 1,\ w_2 = 3,\ w_3 = 5 \longrightarrow p_{3,2} = \frac{5}{5+3} = \frac{5}{8}$

Chess, Xbox ratings

- Fit BTL model, normalize

Also called Plackett Luce [59]

- Pairwise comparisons → ordering of ≥ 2 elements

Parameters are real values

Corrupted by additive Gaussian noise with 0-mean, known-variance

P(12 y/o taller than 10 y/o)

Height at age i — Mean μ_i plus Gaussian noise $X \sim N(0, 1)$

$$
\begin{aligned}
p_{i,j} &= f(\mu_i, \mu_j) \\
&\stackrel{\text{def}}{=} \Pr(\mu_i + X' > \mu_j + X'') \\
&= \Pr(X' - X'' > \mu_j - \mu_i) \\
&= \operatorname{erfc}\left(\frac{\mu_j - \mu_i}{\sqrt{2}}\right)
\end{aligned}
$$

Can have $N(0, \sigma_i)$ for age i

$p_{i,j} \approx \frac{1}{2}$ → many comparisons

BTL

$n = 2$

$(w_1, w_2) \in \left\{(1, 1 + 10^{-10}), (1 + 10^{-10}, 1)\right\}$

$N_{\mathsf{max}} = ?$

$\approx \infty$

PAC Maxing & Ranking [Val 84, Yue et al. 11, Szö et al. 15]

Probably Approximately Correct

a is ϵ-preferable to b: $p_{a,b} \geq \frac{1}{2} - \epsilon$

- Preferable, or "almost preferable"
- $\epsilon = 0.02$, $p_{a,b} = .6$ ✓, $.49$ ✓
- Tiny difference → ¯_(ツ)_/¯

ϵ-maximum: element m that is ϵ-preferable to all others

$$p_{m,x} \geq \frac{1}{2} - \epsilon \quad \forall x$$

ϵ-ranking: permutation $r_1, r_2, ..., r_n$ s.t. whenever $r_i > r_j$,

$$p_{i,j} \geq \frac{1}{2} - \epsilon$$

As before: Answer (for ranking – all) correct with probability $\geq 1 - \delta$

Previous Results [Szö, Busa, Paul, Hull, '15]

For Bradley, Terry, Luce

$$\forall \epsilon, \delta \quad N_{\mathsf{max}} = \mathcal{O}\left(\frac{n}{\epsilon^2} \log \frac{n}{\epsilon\delta}\right)$$

$$N_{\mathsf{sort}} = \mathcal{O}\left(\frac{n \log n}{\epsilon^2} \log \frac{n}{\epsilon\delta}\right)$$

+: Applies to all parameters

−: Factor $\log n >$ deterministic

BTL still too rigid

Arbitrary $p_{i,j}, 1 \le i < j \le n$

$\binom{n}{2}$ parameters

Can this be even done?

For 0.1 maximum: one element beats all others w.p. ≥ 0.4

Define: $\{p_{i,j}\}$ $\longleftrightarrow$ story $\longleftrightarrow$ child's play

$\{$rock, paper, scissors$\}$

$p_{\text{rock,paper}} = 0$ $p_{\text{paper,scissors}} = 0$ $p_{\text{scissors,rock}} = 0$

Everything loses to something with probability 1

Nothing beats everything with probability ≥ 0.4

No 0.1 maximum

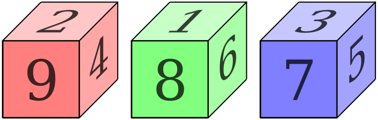
$\epsilon < \frac{1}{2} \longrightarrow$ no ϵ-maximum

Similarly, no ϵ-ranking

Reason: rock < paper < scissors < rock

Non-transitive

No maximum or ranking :-(



(a) (b) (c)

$$P(a > b) = P\{(9,8),(9,6),(9,1),(4,1),(2,1)\} = \frac{5}{9} > \frac{1}{2}$$

$$P(b > c) = P\{(8,7),(8,5),(8,3),(6,5),(6,3)\} = \frac{5}{9} > \frac{1}{2}$$

$$P(c > a) = P\{(7,4),(7,2),(5,4),(5,2),(3,2)\} = \frac{5}{9} > \frac{1}{2}$$

Intransitive dice

Warren Buffet's favorite ice breaker, till met Bill Gates

To even have maximum & ranking $\rightarrow$ need "Transitivity"

If "a beats b" and "b beats c" then "a beats c"

Long studied by behavioral psychologists

Satisfied by standard models (BTL, Gaussian)

Three types

Start with strong

Successively relax

See results we can get

How far we can go

$$p_{a,b} \geq \frac{1}{2}, \quad p_{b,c} \geq \frac{1}{2} \implies p_{a,c} \geq \max(p_{a,b}, p_{b,c})$$

$$p_{\text{Spain, Germany}} = 0.7 \quad p_{\text{Germany, Italy}} = 0.6 \implies p_{\text{Spain, Italy}} \geq 0.7$$

$p_{a,b} \geq \frac{1}{2}, \quad p_{b,c} \geq \frac{1}{2}$ necessary

$$p_{\text{Germany, Italy}} = 0.6 \quad p_{\text{Italy, England}} = 0.1 \not\Longrightarrow p_{\text{Germany, England}} \geq 0.6$$

Transitivity $\longrightarrow$ order $\longrightarrow$ maximum and ranking

$\forall \; w_a \geq w_b \geq w_c$

$$\frac{w_a}{w_a + w_c} \geq \max\{\frac{w_a}{w_a + w_b}, \frac{w_b}{w_b + w_c}\}$$

$$\tfrac{w_a}{w_a+w_c} \geq \tfrac{w_a}{w_a+w_b} \quad \leftarrow \quad w_c \leq w_b$$

$$\tfrac{w_a}{w_a+w_c} \geq \tfrac{w_b}{w_b+w_c} \quad \leftrightarrow \quad \tfrac{1}{1+w_c/w_a} \geq \tfrac{1}{1+w_c/w_b} \quad \leftarrow \quad w_a \geq w_b$$

BTL transitive

Similarly for additive Gaussian Noise

Stochastic transitivity lower bounds $p_{a,c}$: $\quad p_{a,c} \geq \max(p_{a,b}, p_{b,c})$

Sometimes an upper bound is useful too

Natural upper bound: triangle inequality $p_{a,c} \leq p_{a,b} + p_{b,c}$

Trivial for $p_{a,b} \geq \frac{1}{2}, \quad p_{b,c} \geq \frac{1}{2}$

Need more relevant formulation

$\tilde{p}_{i,j} \stackrel{\text{def}}{=} p_{i,j} - \frac{1}{2}$

Reformulations

Anti-symmetry: $p_{b,a} = 1 - p_{a,b}$ $\quad 0.6 \to 0.4$

$\tilde{p}_{b,a} = -\tilde{p}_{a,b}$ $\quad 0.1 \to -0.1$

ϵ-preferable: $p_{a,b} \geq \frac{1}{2} - \epsilon$

$\tilde{p}_{a,b} \geq -\epsilon$

Transitivity: $p_{a,b}, p_{b,c} \geq \frac{1}{2} \to p_{a,c} \geq \max\{p_{a,b}, p_{b,c}\}$

$\tilde{p}_{a,b}, \tilde{p}_{b,c} \geq 0 \to \tilde{p}_{a,c} \geq \max\{\tilde{p}_{a,b}, \tilde{p}_{b,c}\}$

Stochastic Triangle Inequality (STI)

$\tilde{p}_{a,b}, \tilde{p}_{b,c} \geq 0 \implies \tilde{p}_{a,c} \leq \tilde{p}_{a,b} + \tilde{p}_{b,c}$

- $|\tilde{p}_{i,j}|$ – distance between i and j

Parametric models

- Convexity: $x \geq y \longrightarrow \frac{\partial^2 p_{x,y}}{\partial x \partial y} \geq 0$ satisfy STI
- BTL, additive Gaussian models satisfy STI

With ST and TI $N_{\mathsf{max}} = O\left(\frac{n}{\epsilon^2}\log\frac{n}{\epsilon\delta}\right)$

+: Broader model than BTL

−: Same $n \log n$ as BTL

No results for ranking (BTL: $n \log^2 n$)

Max: $n \log n$? n?   in between?

Deterministic: sequential elimination, $n - 1$ comparisons

How simple for randomized?

1 page?   10 pages?   50 pages?

1 page

But... 10 slides

Bias ϵ, allowed error δ

Compare(a, b, ϵ, δ)

- Compares a and b for $\frac{1}{2\epsilon^2} \log \frac{2}{\delta}$ times
- Element winning more times declared match winner
- If $|\tilde{p}(a,b)| \geq \epsilon$, better element wins w.p. $\geq 1 - \delta$

Observations

- $\approx \frac{1}{\epsilon^2} \log \frac{1}{\delta}$ comparisons (next slide)
- Inversely proportional to ϵ and δ
- Algorithm stops earlier if $\tilde{p}(a,b) \gg \epsilon$

\# comparisons $\frac{1}{\epsilon^2} \log \frac{1}{\delta}$ used frequently

Where from?

$n = 2$

$X_1, X_2, \ldots, X_t \sim B(\frac{1}{2} - \epsilon)$ independent

$\overline{X} = \frac{X_1 + X_2 + \ldots + X_t}{t}$

Chernoff bound: $P(\overline{X} \geq \frac{1}{2}) \leq e^{-\frac{t\epsilon^2}{3}}$

Want $\delta \leq e^{-\frac{t\epsilon^2}{3}}$

$\frac{t\epsilon^2}{3} \geq \log \frac{1}{\delta}$

$t \geq \frac{3}{\epsilon^2} \log \frac{1}{\delta}$

Start with algorithm for fixed noise (different from Feige et al.)

First attempt : Sequential Elimination

Pick first element as running element r

Compare r against any remaining element

Update r with winner and remove the other element

Maximum element may be compared to $n-1$ elements

For union bound, each comparison fails w.p. $\leq \frac{\delta}{n}$

comparisons per element: $\mathcal{O}(\frac{1}{\epsilon^2}\log\frac{n}{\delta})$

Total # comparisons: $\mathcal{O}(\frac{n}{\epsilon^2}\log\frac{n}{\delta})$

Ensured *all* comparisons correct

n comparisons, each with error δ/n

Care only about comparisons involving max

In sequential elimination max may be compared $n-1$ times

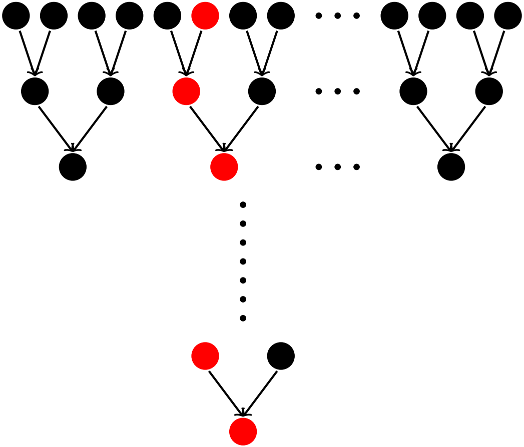
Can we compare max fewer times?

Knockout!

Every element, including max, compared just $\log n$ times

Knockout

$\log n$ levels, compare a pair each time



P_{error}

$\delta / \log n$

$\delta / \log n$

$\delta / \log n$

$\vdots$

$\delta / \log n$

$\vdots$

$\delta / \log n$

Error

By union bound $\frac{\delta}{\log n} \cdot \log n = \delta$

Comparisons

$$n \cdot \frac{1}{\epsilon^2} \log \frac{\log n}{\delta} = \frac{1}{\epsilon^2} \cdot n \cdot \left(\log \log n + \log \frac{1}{\delta} \right)$$

Reduced factor from $\log n$ to $\log \log n$

Can we do better?

Error δ is sum of error δ_i in each stage i

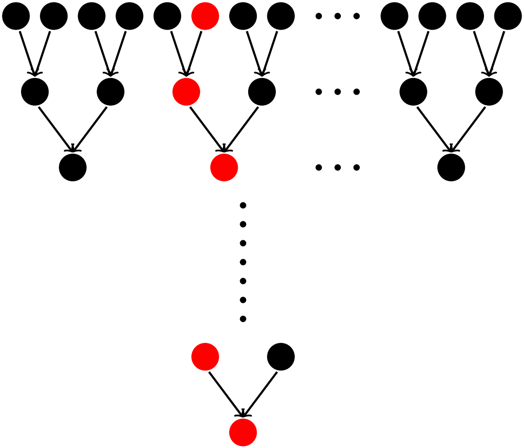
Early stages performed for many pairs

Make δ_i large for early stages, small later, keeping sum same

$\delta_i = \frac{1}{2^i}$

Knockout with Decreasing Error

Knockout: $\log n$ levels, Compare each time



error	new error
$\delta/\log n$	$\delta/2$
$\delta/\log n$	$\delta/4$
$\delta/\log n$	$\delta/8$
$\vdots$	$\vdots$
$\delta/\log n$	$\delta/2^i$
$\vdots$	$\vdots$
$\delta/\log n$	δ/n

Error Probability

By union bound $\sum_{i=1}^{\log n} \frac{\delta}{2^i} \leq \delta$

Comparisons

$$\sum_{i=1}^{\log n} \frac{n}{2^i} \cdot \frac{1}{\epsilon^2} \log \frac{2^i}{\delta} = \frac{n}{\epsilon^2} \sum_i \frac{1}{2^i} \cdot \left(i + \log \frac{1}{\delta} \right) = \mathcal{O}\left(\frac{n}{\epsilon^2} \log \frac{1}{\delta} \right)$$

As easy as flipping a coin

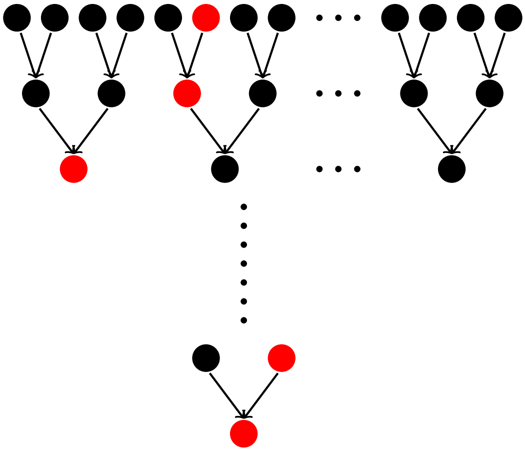
Assumed all $|\tilde{p}_{i,j}| \geq \epsilon$

Any error has small probability

Some $\tilde{p}_{i,j}$ may be smaller than ϵ

Errors may occur, and accumulate

Maxing: ST +STI



bias	confidence
$c\epsilon/2^{1/3}$	$\delta/2$
$c\epsilon/2^{2/3}$	$\delta/4$
$c\epsilon/2$	$\delta/8$
$\cdot$	$\cdot$
$c\epsilon/2^{i/3}$	$\delta/2^i$
$\cdot$	$\cdot$
$c\epsilon/n^{1/3}$	δ/n

$p_{i,j}$ can be $\approx \frac{1}{2}$, $\max$ may lose

Show two properties

ST : Maxim m, m' in two successive stages satisfy $\tilde{p}_{m,m'} < \epsilon_i$

- If m wins, $\tilde{p}_{m,m'} = 0$
- If m loses to m', # comparisons ensures $\tilde{p}_{m,m'} < \epsilon_i$
- If m loses to m'', $\tilde{p}_{m,m'} \le \tilde{p}_{m,m''} < \epsilon_i$

STI: Maximum at beginning and end satisfy $|\tilde{p}_{m,m'}| < \sum \epsilon_i$

With ST and TI

$N_{\mathsf{max}} = \Theta\left(\frac{n}{\epsilon^2}\log\frac{1}{\delta}\right)$

For $\delta > \frac{1}{n}$, $N_{\mathsf{sort}} = \Theta\left(\frac{n\log n}{\epsilon^2}\right)$

+ : More general and stronger than BTL

  Stronger than previous for ST and TI

  Same as non-noisy comparisons

– : Strong transitivity too strong?

$p_{a,b} \geq \frac{1}{2}, \quad p_{b,c} \geq \frac{1}{2} \quad \implies p_{a,c}$

Strong Stochastic Transitivity(SST):

$$\geq \max(p_{a,b}, p_{b,c})$$

Medium Stochastic Transitivity(MST):

$$\geq \min(p_{a,b}, p_{b,c})$$

Weak Stochastic Transitivity(WST):

$$\geq \frac{1}{2}$$

SST $\implies$ MST $\implies$ WST

New Results 2

Previous Results

Model	Maxing	Ranking
Constant Noise	$\mathcal{O}\left(\frac{n}{\epsilon^2}\log\frac{1}{\delta}\right)$	$\mathcal{O}\left(\frac{n\log n}{\epsilon^2}\right)^*$
BTL	$\mathcal{O}\left(\frac{n}{\epsilon^2}\log\frac{n}{\epsilon\delta}\right)$	$\mathcal{O}\left(\frac{n\log n}{\epsilon^2}\log\frac{n}{\epsilon\delta}\right)$
ST with TI	$\mathcal{O}\left(\frac{n}{\epsilon^2}\log\frac{n}{\epsilon\delta}\right)$	—

New Results

<table>
<tr><th>Model</th><th>Maxing</th><th>Ranking</th></tr>
<tr><td>WT with/out TI</td><td>$\Omega(n^2)$</td><td rowspan="3">$\Omega(n^2)$</td></tr>
<tr><td>MT with/out TI</td><td rowspan="3">$\Theta\left(\frac{n}{\epsilon^2}\log\frac{1}{\delta}\right)$</td></tr>
<tr><td>ST without TI</td></tr>
<tr><td>ST with TI</td><td>$\Theta\left(\frac{n\log n}{\epsilon^2}\right)^*$</td></tr>
</table>

* $\delta > 1/n$

Lower Bound For Weak Transitivity

$\Omega(n^2)$ lower bound

Simple question

Mathematical

A...

n pieces in a line

Check two pieces, in order

Find if they match or not

comparisons to assemble the puzzle?

$O(n^2)$

Even just to find the first piece

$\Omega(n^2)$

Connection to maxing

$p_{i,i+1} = 1$, $p_{i,i-1} = 0$, other $p_{i,j} = \frac{1}{2}$

Weakly transitive , not medium or strong

What if no transitivity holds??

Maximum and Ranking can still be defined

Borda scores

$s(e) \stackrel{\text{def}}{=} \Pr(e \text{ is preferable to random element}) = \frac{1}{n} \sum_i p(e,i)$

Maximum and sorting based on these scores

$N_{\mathsf{max}} = \Theta\left(\frac{n}{\epsilon^2} \log \frac{1}{\delta}\right)$

$N_{\mathsf{sort}} = \Theta\left(\frac{n}{\epsilon^2} \log \frac{n}{\delta}\right)$

Approximating all $\binom{n}{2}$ pairs using $\mathcal{O}(n \log n)$ comparisons?

If no SST+STI, ranking $\Omega(n^2)$, hence all $p_{i,j}$ also $\Omega(n^2)$

For SST+STI

Sorting: $\Theta\Big(\frac{n \log n}{\epsilon^2}\Big)$ for $\delta \geq 1/n$

Can we find all $p_{i,j}$ using slightly more comparisons?

Yes. $\Theta\Big(\frac{n \min(n, 1/\epsilon)}{\epsilon^2} \log n\Big)$ comparisons for $\delta \geq 1/n$

$\mathcal{O}(n \log n)$ for $n \geq 1/\epsilon$

[1] Feige, Uriel, Raghavan, Prabhakar, Peleg, David, and Upfal, Eli. Computing with noisy information. SIAM Journal on Computing, 23(5):1001–1018, 1994.

[2] Szörényi, Balázs , Busa-Fekete, Róbert, Paul, Adil, and Hüllermeier, Eyke. Online rank elicitation for plackett-luce: A dueling bandits approach. In NIPS, pp. 604–612, 2015.

[3] Yue, Yisong and Joachims, Thorsten. Beat the mean bandit. In Proc. of the ICML, pp. 241–248, 2011.

[4] Falahatgar, Moein, Orlitsky, Alon, Pichapati, Venkatadheeraj, and Suresh, Ananda Theertha. Maximum selec- tion and ranking under noisy comparisons. In Interna- tional Conference on Machine Learning, pp. 1088–1096, 2017.

[5] Falahatgar, Moein, Hao, Yi, Orlitsky, Alon, Pichapati, Venkatadheeraj, and Ravindrakumar, Vaishakh. Max- ing and ranking with few assumptions. In Advances in Neural Information Processing Systems, pp. 7063–7073, 2017.

[6] Falahatgar, Moein, Jain, Ayush, Orlitsky, Alon, Pichapati, Venkatadheeraj, and Ravindrakumar, Vaishakh. "The Limits of Maxing, Ranking, and Preference Learning." to appear in International Conference on Machine Learning, 2018.

CAREERS

Told To Do Around 10 Interviews, A Job Seeker Pulls Out Of The Running And Shares His Frustration In A Viral LinkedIn Post

Jack Kelly Senior Contributor ⓘ
I write actionable interview, career and salary advice.

Follow

Jun 29, 2021, 03:48pm EDT

Thank You!