Strategic ranking

Lydia T. Liu[1], Nikhil Garg[1,2] and Christian Borgs[1]
[1]University of California, Berkeley [2]Cornell Tech and Technion

Background

Many consequential decisions are based on *relative*, not absolute, measures of quality

- The literature on algorithmic fairness and strategic behavior [e.g. 1, 2, 3] has focused on classification; constrained allocation and ranking (e.g. college admissions) has received little attention.

Strategic and fairness considerations are relevant in the design of rankings, but not well understood

- Strategic individuals may exert effort to influence their rankings, depending on rewards.
- Different groups of individuals may have different returns to effort.

Main Contributions

- **We study *strategic ranking*, where an applicant's reward depends on their post-effort rank in a measured score.**
- We illustrate the **equilibrium behavior that results from competition among applicants**, and show how ranking reward designs **trade off applicant, school, and societal utility.**
- **Randomization** in the ranking reward design can **mitigate** two measures of **disparate impact: welfare gap and access.**

Model

Applicants Unit mass, indexed by $\omega \in [0,1]$ distributed uniformly.

Latent skill rank. unobserved $\theta_{\text{pre}} = \theta_{\text{pre}}(\omega) \in [0,1]$. Skill of applicant is $f(\theta_{\text{pre}})$, where f strictly increasing, continuous.

Effort and Score. Applicant chooses effort level $e \geq 0$. The result is an observed, post-effort *score*, $v = v(e, \theta_{\text{pre}}) = g(e) \cdot f(\theta_{\text{pre}})$. The effort transfer function g is continuous, concave, strictly increasing (marginal effort improves one's score but has diminishing returns).

References

[1] Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, ITCS '16, pages 111–122, New York, NY, USA, 2016. ACM.

[2] Lily Hu, Nicole Immorlica, and Jennifer Wortman Vaughan. The disparate effects of strategic manipulation. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 259–268, New York, NY, USA, 2019. ACM.

[3] Smitha Milli, John Miller, Anca D. Dragan, and Moritz Hardt. The social cost of strategic classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 230–239, New York, NY, USA, 2019. ACM.

[4] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

[5] Wojciech Olszewski and Ron Siegel. Large contests. *Econometrica*, 84(2):835–854, 2016.

Model - continued

Post effort rank. Each applicant is ranked according to their score v , resulting in a *post-effort rank* θ_{post}.

School. Admits applicants, according to ranking reward function $\lambda : [0,1] \mapsto [0,1]$, s.t. an applicant with post-effort rank θ_{post} is admitted with probability $\lambda(\theta_{\text{post}})$. λ is non-decreasing and the school has a capacity constraint, i.e., $\mathbb{E}[\lambda(\theta_{\text{post}})] = \rho$.

Individual applicant welfare. Given the designer's function λ and the effort levels of other applicants, each applicant chooses effort e to maximize their individual welfare,

$$W(e, \lambda(\theta_{\text{post}})) = \lambda(\theta_{\text{post}}) - p(e).$$

where p is non-negative, continuous, and strictly convex.

Equilibrium. After a school chooses its ranking reward function λ, an *equilibrium* of effort levels is an assignment $\theta_{\text{pre}} \mapsto e(\theta_{\text{pre}})$ of effort levels and resulting post-effort ranks $\theta_{\text{post}}(\theta_{\text{pre}}(\omega))$ in which given the efforts of other applicants, no applicant can increase their welfare by changing their effort.

Equilibrium characterization

- **Rank preservation**: In every equilibrium, $\lambda(\theta_{\text{post}}(\theta_{\text{pre}})) = \lambda(\theta_{\text{pre}})$, up to sets of measure 0.
- **Second price effort**: Each applicant exerts just enough effort that applicants in the level below (of pre-effort rank) cannot increase welfare by exerting additional effort (cf. [4] and [5])

Main Results

Tradeoffs in aggregate welfare and utility

Applicant welfare. $\mathcal{W} := \mathbb{E}[W(e, \lambda(\theta_{\text{post}}))] = \rho - \mathbb{E}[p(e)]$

Societal utility. $\mathcal{U}^{\text{soc}} := \mathbb{E}[v]$

Private utility. $\mathcal{U}^{\text{pri}} := \mathbb{E}[v \cdot \lambda(\theta_{\text{post}})]$

Two-level policy. Parameterized by cut-off $c \in (0, 1-\rho]$, an applicant with post-effort rank $\theta_{\text{post}} \geq c$ is admitted with probability $\ell_1 = \frac{\rho}{1-c} > 0$. All others are rejected. **Lower c = more randomized admissions.**

Result. In the class of two-level policies, $\mathcal{W}$ is decreasing with c. $\mathcal{U}^{\text{pri}}$ is increasing with c. $\mathcal{U}^{\text{soc}}$ may be maximized at $c \in (0, 1-\rho)$.

Environment difference and structural inequality

Suppose there are now two groups, A, B with environment factors $\gamma_{\mathsf{A}} > \gamma_{\mathsf{B}} > 0$. **Favorable environment results in higher return to effort**: $v = \gamma \cdot g(e) \cdot f(\theta_{\text{pre}})$

Let $\mathcal{W}^{\mathsf{G}}(\theta_{\text{pre}})$ denote post-effort welfare of a applicant with latent skill ranking θ_{pre} from group $\mathsf{G} \in \{\mathsf{A}, \mathsf{B}\}$, i.e.,

$$\mathcal{W}^{\mathsf{G}}(\theta_{\text{pre}}) := \lambda(\theta_{\text{post}}(\theta_{\text{pre}}, \gamma_{\mathsf{G}})) - p(e(\theta_{\text{pre}}, \gamma_{\mathsf{G}})).$$

Welfare gap. $\mathcal{G}(\theta_{\text{pre}}) := \mathcal{W}^{\mathsf{A}}(\theta_{\text{pre}}) - \mathcal{W}^{\mathsf{B}}(\theta_{\text{pre}})$.

Result. In the class of two-level policies, $\mathcal{G}(\theta_{\text{pre}})$ is strictly decreasing with c for all θ_{pre} above a threshold.