Data Driven Optimization in Market Design

Justin Payan

December 15, 2023

I apply techniques from combinatorial optimization, market design, and data science to high-stakes social problems. My main application is peer review, but I have applied my skills to other problems. Because I focus on problems of social importance, I am very interested in translating my algorithms into practice. The driving questions of my research are: *What outcomes do practioners care about, and how can we ensure those outcomes are achieved?*, and *What are the barriers to applying existing solutions in novel settings?*. I will illustrate how each of these questions has driven my prior and current work in peer review, and how I hope to answer these questions during the next phase of my academic career.

1 FairSequence

Our first contribution to peer review is the FairSequence reviewer assignment algorithm, which has been deployed in the major conference management platform OpenReview.[1] It is based on preliminary work published at IJCAI 2022 [6]; a more complete analysis is under submission to JAIR. FairSequence assigns reviewers to papers fairly, efficiently, and quickly while remaining flexible to the constantly evolving constraint set required by conference organizers.

We initially developed the *Greedy Reviewer Round Robin* (`GRRR`) algorithm. Like most reviewer assignment algorithms, `GRRR` takes as input *affinity scores*, which are real numbers measuring the suitability of each reviewer to review each paper. Our objective is to assign each paper p a set (or bundle) of reviewers A_p. The utility of each paper for their assigned bundle of reviewers is measured via a *utility function* $v_p(A_p)$. `GRRR` ensures that papers are *envy-free up to one item*, or EF1: papers prefer their assigned reviewers to those assigned to other papers, or at least do so with the exception of one reviewer. More formally, the paper p is EF1 with respect to the paper p' if either $v_p(A_p) \geq v_p(A_{p'})$ or there is some reviewer $r \in A_{p'}$ such that $v_p(A_p) \geq v_p(A_{p'} - r)$. We can achieve this criterion in our assignments by ordering papers in some order $\mathcal{O}$ and repeatedly assigning the best reviewer to each paper in that order. However, an arbitrary order may not maximize the total affinity between reviewers and papers. Our main technical result shows that greedily maximizing the order approximately optimizes welfare under any round-robin picking sequence.

Unfortunately, we found that greedily selecting a picking sequence for `GRRR` took a matter of days on medium size conferences (such as CVPR 2018, with about 5000 submissions and 3000 reviewers). Conference organizers often run algorithms multiple times, and they must decide on a final allocation within a matter of days. In addition, after approaching the OpenReview team with `GRRR`, they informed us that any algorithm they deploy must adequately handle non-uniform paper demands,

[1] `https://github.com/openreview/openreview-matcher`

and must be able to handle lower bounds for reviewers. The twin concerns of computational efficiency and handling those 2 practical constraints lead us to develop `FairSequence`. `FairSequence` uses a modified picking sequence to guarantee *weighted* EF1, a variation of EF1 that applies when papers have varying demands. We also modify the algorithm to be much faster, showing nearly an order of magnitude speed-up compared to all existing reviewer assignment approaches. We show that `FairSequence` has a much lower Gini coefficient than OpenReview's previous state-of-the-art fair assignment algorithm, suggesting a deeper connection between envy-based constraints and traditional global fairness measures.

I hope to further study the deeper connections between envy and Gini inequality, showing stronger theoretical justification for usage of envy-based fairness notions in more applications. There are other constraints that are increasingly incorporated into reviewer assignment, such as aiming to have a wide range of seniority levels and geographic origins among the set of reviewers assigned to each paper [3]. Incorporating these constraints into `FairSequence` will allow many more conferences to benefit from its speed and fairness.

2 Robust Reviewer Assignment

While developing `FairSequence`, we often confronted the question *How do we know if a reviewer is a good fit for a paper?* Our framework Robust Reviewer Assignment (RRA) assigns reviewers to conference papers by building an uncertainty-aware, predictive model of review quality and ensuring high quality reviews are likely [1]. RRA first constructs a (δ, γ)-*uncertainty set* $\mathcal{S}$, such that the true affinity score matrix S^* is γ close to a point in $\mathcal{S}$ with probability $1 - \delta$. We then select an allocation A to optimize

$$\max_{A} \min_{S \in \mathcal{S}} \mathrm{USW}(A, S) \ . \tag{1}$$

Our paper outlines theoretical approaches to construct uncertainty sets $\mathcal{S}$ from historical conference data or partially-observed data from ongoing conferences, and we provide an algorithm to approximately solve 1 for relevant instantiations of $\mathcal{S}$.

There are still many open questions in this line of work. The robust welfare maximization problem 1 does not incorporate fairness. Although we cannot robustly guarantee maximum egalitarian welfare [2, 7], we are currently working on providing *group* egalitarian welfare guarantees.

RRA provides a framework for estimating uncertainty sets $\mathcal{S}$ from data and solving 1 over $\mathcal{S}$. However, we still have to describe in detail how to construct $\mathcal{S}$. Often, we have partially-available information from an ongoing conference indicating the quality of reviewer-paper pairs. For example, IJCAI 2023 implemented a two-stage reviewing process where SPCs in the first stage rated the quality of a limited number of reviews. In ongoing work, we propose using this limited rating data to estimate an uncertainty set $\mathcal{S}$ for the remaining pairs. We give a detailed construction of the estimator, along with error bounds to show that the true ratings will lie within γ of $\mathcal{S}$ with high probability.

Although theoretically showing robustness to uncertainty is an important goal, ultimately we cannot ensure that our decisions are good without comparing our objectives to outcomes of interest in real conferences. I served as Workflow Chair of IJCAI 2023, and am in talks to serve again for IJCAI 2024. In this capacity, I am working with a team to further analyze the data from last year's conference to understand how variables available during reviewer assignment (e.g., reviewer bids, document similarity measures, keyword matching) correlate with and predict measures of review quality (SPC ratings of review quality, reviewer confidence, and fine-grained annotations of review texts).

3 Applications in Novel Settings

Our reviewer assignment algorithms today assume that reviewers are assigned to a batch of submitted conference papers all at once, and that the assignment produced by the allocation algorithm is the final assignment deployed by the conference organizers. For reviewer assignment algorithms to realize their full potential, we need to make them much more flexible to alternative deployment contexts.

One important overlooked context is journal and rolling review scenarios.[2] In this setting, we have a pool of reviewers that serve on a program committee over an extended period of time. These reviewers typically have maximum workloads for any given point in time. In addition, if they are assigned the maximum workload every time step, they will become burnt out and decline further reviewing requests. Thus, we must make an immutable assignment of a subset of reviewers at each time step, without depleting the reserves of reviewers that will be required in the future. We are currently designing an algorithm that samples future submissions and maximizes the average affinity over samples. We can then provide high probability bounds on the regret of this algorithm.

I am also interested in supporting standard two-stage reviewing formats. In this context, we might ask *How do we identify papers that require additional reviewers?* Once we answer that question, we must also define and satisfy reasonable efficiency and fairness notions when part of the allocation is already fixed (i.e., the first-stage allocation is fixed, and we must ensure efficiency and fairness in the second stage).

4 Broader Goals

My broader goal is to use market design and data science to improve peer review and other application areas that are typically neglected by industry. Although I have focused on combinatorial optimization for reviewer assignment, I am interested in further expanding my toolbox in the next several years. For instance, I recently asked a program chair of a major conference what the most impactful research I could do is; their reply was to understand the interplay between conference submission rates, prestige, and size, and suggest changes that will naturally guide conferences to shrink in size. Tools from the theory of congestion games might be applied here. I also believe that large language models can have a strong role to play in my future work. I have worked on LLMs with multiple industry research teams [4, 5], and I think that LLMs can be used in market design problems — as a way of collecting preferences, or as an interface for downstream users.

Designing algorithms that are used in practice requires working with practitioners. I have developed useful relationships with OpenReview and the management team for IJCAI, and I hope to build on these relationships so I can continue to address practical hurdles to improvements in peer review. Deploying and monitoring research is challenging, but often leads to new ideas (`FairSequence` and our analysis of the IJCAI 2023 dataset are prime examples of this principle). I intend to build this muscle throughout my academic career, so I can ensure my work has maximum impact.

[2] ACL Rolling Review is a highly successful example: `https://aclrollingreview.org/`

References

[1] C. Cousins, J. Payan, and Y. Zick. Into the unknown: Assigning reviewers to papers with uncertain affinities. In *Proceedings of the 16th International Symposium on Algorithmic Game Theory (SAGT)*, 2023.

[2] A. Kobren, B. Saha, and A. McCallum. Paper matching with local fairness constraints. In *Proceedings of the 25th International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1247–1257, 2019.

[3] K. Leyton-Brown, Y. Nandwani, H. Zarkoob, C. Cameron, N. Newman, D. Raghu, et al. Matching papers and reviewers at large conferences, 2022.

[4] J. Payan, Y. Merhav, H. Xie, S. Krishna, A. Ramakrishna, M. Sridhar, and R. Gupta. Towards realistic single-task continuous learning research for ner. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3773–3783, 2021.

[5] J. Payan, S. Mishra, M. Singh, C. Negreanu, C. Poelitz, C. Baral, S. Roy, R. Chakravarthy, B. Van Durme, and E. Nouri. Instructexcel: A benchmark for natural language instruction in excel. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023.

[6] J. Payan and Y. Zick. I will have order! Optimizing orders for fair reviewer assignment. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*, 2022.

[7] I. Stelmakh, N. B. Shah, and A. Singh. PeerReview4All: Fair and accurate reviewer assignment in peer review. In *Proceedings of the 30th International Conference on Algorithmic Learning Theory (ALT)*, pages 828–856, 2019.