Implied Ordinal Preferences

Ian Shaw
ianedwardshaw@gmail.com

May 4, 2019

Abstract

This paper investigates the the possibility of uncovering implied ordinal preferences when provided incomplete ordinal preferences for one side of a matching market. The work of Gale and Shapely's [8] provided the first algorithm for solving the stable marriage and attempted the more general college admissions problems (verified as different problems by Roth [13]). Much of the work in this field focuses on stability, which is not necessarily a requirement in systems where market participants are compelled [12]. In situations of compulsion, the preference u (un-assigned) is not an option. This arises in legally compulsory assignment situations such as jobs in the military or secondary school enrollment.

In these situation, complete ordinal preferences are not always provided. This can be due to a lack of time, lack of knowledge about preferences, excess of options, or some other system deficiency that does not encourage/enable participants to express preferences on all matching options.

This paper proposes a system where, given at least one preference for all participants on one side of the market, preferences may be supposed at a higher accuracy than random preference assumption.

The code to demonstrate the matching algorithms, optimization, and preference-based metrics can be found in Ian Shaw's Github Repository. [1]

[1] https://github.com/ieshaw/Imp_Ord_Pref

Contents

1 Formulation

Consider a worker i and an opportunity at a firm j. Suppose there are n firms, but worker i not expressed complete preferences, $n_i < n$. Suppose there are m workers in the compulsory market.

Consider the following terms

$$\begin{aligned}
J &= \text{set of opportunities at firms in the market} \\
|J| &= n \\
P_i &\in \mathbb{Z}^{+, n \times 1} \\
P_{ij} &= \begin{cases} p & p \in \mathbb{Z}^+ > 0, \text{expressed ordinal preference of worker } i \text{ for job } j \\ 0 & \text{worker } i \text{ has not expressed an ordinal preference for job } j \end{cases} \\
P_i^C &\in \mathbb{Z}^{+, n \times 1} \\
P_{ij}^C &= \begin{cases} n - p & p \in \mathbb{Z}^+ > 0, \text{expressed ordinal preference of worker } i \text{ for job } j \\ 0 & \text{worker } i \text{ has not expressed an ordinal preference for job } j \end{cases} \\
J_i &= \{j : P_{ij} \neq 0\} \\
|J_i| &= n_i \\
J_i' &= J \setminus J_i \\
|J_i'| &= n - n_i \\
S_{i,i'} &= \text{similarity of worker } i \text{ and } i'
\end{aligned}$$

Our proposed implied ordinal preference system P_i' is developed in the following manner. The metric r is the similarity of the worker in question with another worker multiplied by the complement ordinal ranking of that worker (so that metric increases as similar workers more prefer the positions in question), summed across all workers. The metrics are then sorted in descending order. Ties are broken randomly. Then the ordinal ranking of these metrics are considered the implied ordinal preference for the for previously unexpressed preferences.

$$\begin{aligned}
r_{ij} &= \sum_{i'}^{m-1} S_{i,i'} P_{i',j}^C \\
R_i &= [r_{ij} : P_{ij} == 0] \\
R_i[k] &\geq R_i[k+1] \\
P_i' &\in \mathbb{Z}^{+, n \times 1} \\
P_{i,j}' &= \begin{cases} p & p \in \mathbb{Z}^+ > 0, \text{ expressed ordinal preference of worker } i \text{ for job } j \\ n_i + k & R_i[k] == r_{ij}, \text{ implied ordinal preference of worker } i \text{ for job } j \end{cases}
\end{aligned}$$

1.1 Similarity Measures

We explore three types of similarity measure in our investigation: cosine similarity, normed Euclidean distance, and weighted Euclidean distances

Note that we use the complement of the ordinal preference, because having information up to the $n^{th} - 1$ preference is equivalent to having the n^{th} preference; we indicate having no preference information as a preference 0. Thus the most preferred choice is given a value of n.

The cosine similarity of two workers i, i' when considering their preference vectors P, is

$$S_{i,i'} = \frac{P_i^C \bullet P_{i'}^C}{||P_i^C|| \, ||P_{i'}^C||}$$

Normed Euclidean distances is calculated as

$$S_{i,i'} = 1 - \frac{||P_i^C - P_{i'}^C||}{||P_i^C||||P_{i'}^C||}$$

The two measures above are chosen due to their established usage. Below we propose what we call weighted Euclidean distance, intended to punish dissimilar preferences more if highly preferred by one of the workers.

$$S_{i,i'} = 1 - \frac{1}{2m^4}\sum_{k=1}^{m}(P_{i,k}^S + P_{j,k}^S)(P_{i,k}^S - P_{i',k}^S)^2$$

2 Analysis

2.1 Process

Given a set of preferences for one side of the market (does not need to be complete[complete means having preferences for all possibilities]), replace a certain proportion of known preferences with an "unknown" indicator (obviously have to work bottom up because if someone ranks their top 5 and their 7th, then they would know their 6th)[leaving the idea of knowing most and least preferred but not middle preferred for future work]. With each dropout propose two sets of preferences, one with random assignment up to complete preferences, and one using the implied ordinal preference system up to complete preference. Compare the Root-mean-squared error (RMSE) of the two proposed preference sets to the actual preference set.

2.2 Toy Example

Consider the incomplete preferences found in Table 1. Since most of our calculations are done with the complement of the ordinal preferences we also provide these in Table 2.

Table 1: Incomplete preferences. Blank means unexpressed.

Job Option	Seeker 1	Seeker 2	Seeker 3	Seeker 4	Seeker 5
Job A	1	3		1	1
Job B	2	2	1		
Job C	3	1			

Table 2: Complement of incomplete preferences. Unexpressed preferences are replaced with 0.

Job Option	Seeker 1	Seeker 2	Seeker 3	Seeker 4	Seeker 5
Job A	2	0	0	2	1
Job B	1	1	2	0	0
Job C	0	2	0	0	0

Now from here we can calculate the cosine simiilarity of each seeker to each other, this is in Table 3.

We do a matrix multiplication of the complement of the experessed preferences with the similarity scores to get the interim preference score matrix R, found in Table 4.

Finally, we use these scores to complete the preferences in Table 5. As you can see, we have a tie to break for Seeker 3. In this case, we simply choose with rondom uniform probability.

2.3 Data Sources

We were able to gather ordinal preferences on job assignment with the support of the following groups:

If possible, we would be interested in extending our analysis to non-military groups. Particulary, our research has gleaned a possbily beneficial application to the New York and Boston School systems. Another potential source would be the National Residency Match preference data.

Table 3: Cosine Similarity between Job Seekers.

Job Seeker	**Seeker 1**	**Seeker 2**	**Seeker 3**	**Seeker 4**	**Seeker 5**
Seeker 1	1	$\frac{1}{5}$	$\frac{1}{\sqrt{5}}$	$\frac{2}{\sqrt{5}}$	$\frac{2}{\sqrt{5}}$
Seeker 2	$\frac{1}{5}$	1	$\frac{1}{\sqrt{5}}$	0	0
Seeker 3	$\frac{1}{\sqrt{5}}$	$\frac{1}{\sqrt{5}}$	1	0	0
Seeker 4	$\frac{2}{\sqrt{5}}$	0	0	1	1
Seeker 5	$\frac{2}{\sqrt{5}}$	0	0	1	1

Table 4: Toy Matrix R. Known preferences have blanks for scores.

Job Option	**Seeker 1**	**Seeker 2**	**Seeker 3**	**Seeker 4**	**Seeker 5**
Job A			$\frac{2}{\sqrt{5}}$		
Job B				$\frac{2}{\sqrt{5}}$	$\frac{2}{\sqrt{5}}$
Job C			$\frac{2}{\sqrt{5}}$	0	0

Table 5: Completed Preferences.

Job Option	**Seeker 1**	**Seeker 2**	**Seeker 3**	**Seeker 4**	**Seeker 5**
Job A	1	3	3	1	1
Job B	2	2	1	2	2
Job C	3	1	2	3	3

Table 6: Sources of Data Used in the Report.

Data Set	**Participants**	**Options**	**% Complete**
US Navy Medical Corps Doctors	59	20	34.58%
US Navy Medical Corps Hospitals	20	59	13.90%
US Navy Explosive Ordinance Displosal Officers	43	15	100%
US Navy Cryptologic Warfare Officers	28	19	100%
US Navy Cryptologic Warfare Commands	19	28	100%
US Army Officers	108	137	24.15%

2.4 Results

3 Literature Review

3.1 Completed Thus Far

The most important people in this field, and the winners of the 2012 Nobel Prize in Economics for their work in stable marriage matching, are Roth [2] and Shapely [3].

1962

The intent of this algorithm is to provide stable pairings between job owners and job seekers based on their ranked preferences. The algorithm's initial conception and definition of stability can be found in Gale and Shapely's 1962 publication in the January *The American Mathematical Monthly* [8]. The algorithm completes in polynomial time and was originally written for application in collage admissions.

[2] http://stanford.edu/~alroth/PapersPDF.html
[3] http://www.econ.ucla.edu/shapley/ShapleyBiblio.1.html

1982

Roth explores the incentives of conveying true preferences and whether is it in everyone's best interest to do so. [12] In his work he specifically points to the applications to "civil servants with civil service positions".

1985

Roth explores the stable marriage problem specifically in the terms of 'firms and workers', also calling upon the lens of game theory. [14] He discussed an extension of the model from one assignment for each worker or firm, to multiple workers for each firm, to a situation where each firm can have multiple workers and each worker could have multiple firms. He also explored, under the constraint of stability, how in each model the optimal assignment set for one party (eg: firms) is the least optimal for the other (eg: workers). He elaborates that this final phenomenon creates difficulty in the institutional decision of how to formulate the matching algorithm.

1989

Irving explored indifference preferences and the follow on adoption for the Gale-Shapely algorithm. [9] This provides the theoretical framework allowing for indifference in our own formulation. Though much of his focus is on differing forms of stability (weak, strongly, and super) these lie outside of our investigation due to the Navy's authority to compel its members to placement.

1993

Roth, Rothblum, and Vande Vate explored the concept of partial matches, discovering in fact this forms a lattice of solutions as well. These fractional matches could represent lotteries or time splitting. [2]

2001

Mandler executes a deeply technical discussion on the compromises between cardinal and ordinal preference systems in social choice. [11]

2006

Atila et. al elaborate on the investigation into and asubsequent process of changin the Boston school choice mechanism. [**?**] The particular motivation was the lack of strategy proofness of the old mechanism which they overcome by implementing deferred acceptance. Particularly interesting to the implied preference algorithm, is much of the diffuclty comes from unclear advice to parent and student on how to order their school preferences.

2013

Nathanson et. al review the high school choiced and placements, focusing particulalry on low-achieving students. [10] Their investigation notes how student's choices are "shaped and constrained by their familiarity with and proximity to specific schools" which could lead to further socioeconomic segregation. Furthermore, many assignments fall out of the top 5, requiring a second round of preference inquiry, which usually has lower participation. There is also a highly heuristic process of schools handing students, something the IOP algorithm can support in a more continuous, distributed work, approach. "The average applicant ranked between 6 and 7 schools - that is, fewer than the possible 12".

Glaeser et. al review the school preferences expressed by Boston parents. [7] They note that Boston has an open call for additional proposals to improve their assignment system.

References

[1] P. Pathak A. Abdulkadiroglu, A. Roth and Sonmez T. Changing the boston school choice mechanism. Technical report, Boston Public Schools, January 2006.

[2] U. Rothblum A. Roth and J. Vande Vate. Stable matchings, optimal assignments, and linear programming. *Mathematics of Operations Research*, 18(4), November 1993.

[3] S. Board. Preference and utility. UCLA, Department of Economics, October 2009.

[4] G. Carroll. On mechanisms eliciting ordinal preferences. Stanford University, Department of Economics, December 2017.

[5] W. Chiu. An ordinal theory of risk aversion. University of Manchester, January 2017.

[6] F. Kojima P. Milgrom E. Budish, Y. Che. Designing random allocation mechanisms: Theory and applications. *American Economic Review*, 103(2):585–623, 2013. `http://dx.doi.org/10.1257/aer.103.2.585`.

[7] S. Poftak E. Glaeser and K. Tobio. What do parents want? an exploration of school preferences expressed by boston parents. Harvard Kennedy School Faculty Research Working Paper Series, July 2013.

[8] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69:9–15, January 1962.

[9] R. Irving. Stable marriage and indifference. *Discrete Applied Mathematics*, 48:261–272, 1994.

[10] S. Corcoran L. Nathanson and C. Baker-Smith. High school choice in new york city: A report on the school choices and placements of low-achieving students. Technical report, NYU Institute for Education and Social Policy, April 2013.

[11] M. Mandler. Compromises between cardinality and ordinality in preference theory and social choice. Discussion paper, Cowles Foundation for Research in Economics, Yale University Box 208281, New Haven, Connecticut 06520-8281, August 2001. Discussion Paper No. 1322.

[12] A. Roth. The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7(4), November 1982.

[13] A. Roth. The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory*, 36:277–288, 1985.

[14] A. Roth. Conflict and coincidence of interest in job matching: Some new results and open questions. *Mathematics of Operations Research*, 10(3), August 1985.

[15] S. Mitchell S. Khuller and V. Vazirani. On-line algorithms for weighted bipartite matching and stable marriages. *Theoretical Computer Science*, 127:255–267, May Year.

4 Future Work

We hope to extend this work to non-compelling situations, where participants can express preference up to a point, after which their preference is to not participate in the market. Futhermore, a future advancement could be to normalize the preference completion to be independent of certain factors such as race or geography. Furhter research could also be supported by a better experimental error metric than RMSE.