Algorithmic Dynamic Matching
Detailing Marketplace Pilot for US Navy

Ian Shaw, LTJG, USN * Zac Dannelly, LTJG, USN †
Drew Calcagno, LTJG, USN ‡ Richard Kuzma, LTJG, USN §
Brian Yarborough, 1LT, USAF ¶

February 21, 2019

Abstract

In this project we apply Alvin Roth's work in mixed integer programming solution to the college admissions problem to the placement of U.S. Navy sailors to jobs. We further the body of work by providing alternate formulation to the integer program and metrics that are gleaned from the expressed ordered preferences of both the sailors for jobs and the job owners preferences over sailors. These metrics are: specialization, competitiveness, similarity, and preference correlation. Though our presentation is tailored to meet the stated fiscal year 2019 goals of the Chief of Naval Personnel, the work is agnostic and can be applied to any matching market not requiring stability as market participants can be compelled. [1]

* E stuff
† Z stuff
‡ X Stuff
§ Q Stuff
¶ Brian Stuff

[1] The code to demonstrate the matching algorithms, optimization, and preference-based metrics can be found in Ian Shaw's Github Repository `Dynamic_Manning`. `https://github.com/ieshaw/Dynamic_Manning`

Contents

1 Introduction **3**
1.1 Problem Description 3

2 Literature Review **4**

3 Algorithm **5**

4 Metrics **5**
4.1 Pre-Match Metrics 5
4.1.1 Competitiveness 5
4.1.2 Post-Match Metrics 5

5 Pilot **6**
5.1 Overview 6
5.2 Data 6
5.3 Results 6

6 Alternate Formulations **6**
6.1 Co-Location 6
6.1.1 Explanation of Co-location constraint math 7
6.2 Tweaking Objective Function 7
6.3 Wagering 8
6.4 Wagering 8
6.5 Specialization Objective 8

7 Future Work **8**

Appendices **9**

A Notation **9**

B Matching Algorithms **9**

1 Introduction

Under the backdrop of the President of the United States issuing the Executive Order on Maintaining American Leadership in Artificial Intelligence in February 2019, we created the algorithmic foundation for advancement of artificial intelligence and machine learning (AI/ML) in military talent allocation. Before the US Government can accomplish its both lethal and non-lethal AI/ML goals, however, it must embrace the essential building blocks underlying the buzzwords. One of those cornerstones is algorithmic linear programing, specifically the use of mixed integer programming (MIP). In this piece, we reveal the most exquisite MIP algorithm completed thus far for the US Navy.

We believe the world to be arcing toward the need for humans to grow as increasingly specialized workers, and in the vicious games of statecraft and war, allocating talent with hyper specificity will become a necessity for victory. To that end, we theorized, coded, and prototyped an MIP algorithm that matches job seekers (sailors) and job owners (commanding officers).

1.1 Problem Description

Human talent allocation is a resource intensive process for an institution. This paper outlines the mathematics necessary to lay the groundwork for such a task within one of the most complex institutions of the American psyche – the Department of Defense (DoD). We dove into this problem with a specific service branch in mind. Focusing our work on the Navy, and within it, the sailors assigned to the National Security Agency (NSA) for cryptologic and cyber warfare efforts, we honed our test case to a group of officers most aligned with technical advancement. Furthermore, despite the small prototype sample, the algorithm is greatly scalable.

Market forces appear are largely the drivers behind a job marketplace – this concept is not new for the American economy. However, given a captive audience such as service members ordered to specific jobs, the mathematical benefit of benevolent autocracy is immense. Rather than the suboptimal fate to which current servicemembers are subjected, the DoD is shopping for a better way forward.

In the Navy, terminology of a detailing marketplace refers to the attempts to detail, or place, warfighters into their next job. This detailing marketplace is a system by which military members can transparently rank their job preferences and the people who own those jobs can supply their preference of incoming personnel. This problem is rather unique to the military due to the aforementioned captive market of many members obligated to remain in service due to contract, desiring to stay in for a pension, wishing to stay in through a sense of service, complemented by an inability for lateral entry – almost all members needing to start from the entry level. Additionally, servicemembers change jobs nearly every one to three years.

There are ontological parallels to other communities, too. This process is shared in some regards by the medical school graduates applying to the residency stage of their training. These graduates are applying to a pool of US residency programs. This pool is also rather narrow, mostly coming from US based medical schools, and the specificity of skill set required for success is better understood than success as a military officer.

This general similarity is the reason our initial matching algorithm (Gale-Shapely Deferred Acceptance Algorithm) is similar one used by that process, the National Residency Match Program. The Gale-Shapely algorithm's application to the residency matching problem earned the Nobel Prize in 2012.

This paper also proposes an optimization based solution to the matching process. The optimization is not found in the National Residency Match Program as medical students have the ability to reject their assigned position, a choice not always given to military members. Thus an optimization can often more optimal solution for the DoD system at the expense of a few forced members.

Due to the importance of co-locating dual-military households (where two family members are in the military), we focus our matching algorithm and optimization proposals on solutions that would guarantee 95% or greater co-location rate.

Acknowledging the difficulty of wrangling disparate and dated personnel data, this paper also explores helpful metrics that can be gleaned simply from submitted, ordered preferences by job seekers and job owners. These are competitiveness, similarity, generalism, and specialization. Interestingly, due to the fact that preferences are expressed on job seekers and the jobs themselves, these metrics can be developed about the jobs or the job seekers.

Further the paper ends with suggested metrics that would be gleaned if personnel data beyond preferences was accessible, clean, and structured. These include a similarity measure based on quality encodings and a suggested ordering of possible jobs or applicants. The latter is proposed to be enabled by deep learning (the underlying technology of Artificial Intelligence (AI)). The suggested ordering would not make decisions on placement, but rather provide job seekers and job owners with metrics distilling the vast amount of information about

2 Literature Review

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* [2]. The algorithm completes in polynomial time and was originally written for application in collage admissions.

1982

Roth explores the incentives of conveying true preferences and whether is it in everyone's best interest to do so. [4] 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. [5] 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. [3] This provides the theoretical framework allowing for indifference in our own formulation. Though much of his focus is on differing forms of stability (weak, stongl, 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. [1]

1994

Khuller et.al have developed an algorithm for stable matching on-line (matching people as they enter the system), as opposed to the typical formulation of having complete market participants and preferences at the time of matching. [6] This could be interesting in future work of understanding the Navy detailing process as a continuous, rather than discrete, process.

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

3 Algorithm

A basic algorithm, like Gale-Shapelys set of instructions for stable marriages, is an important first step in talent allocation. We moved onto a linear programming concept, oriented around solvers – algorithms that are adaptable. These solvers have varying types like linear, convex, and mixed integer. Our group followed a path paved by Stanford operations researcher Alvin Roth and adapted a mixed integer solution to leverage its flexibility. While Gale-Shapelys algorithm can be solved in n-squared time, ours is n-p-complete, indicating polynomial time completion. What we lose in speed, we gain in the ability to alter the formulation without total overhaul. These qualities make the approach particularly conducive to constraint creation, a vital requirement for stakeholders utilizing our solution.

Ours is a unique subset of mixed integer programming (MIP), binary optimization, to reflect the nature of either placement in a specific job, or not (i.e. only 1 or 0). The resulting lattice has an incredibly large number of dimensions manifesting every potential job placement for every individual. Subsequently, we apply constraints – the inherent reason we utilize MIP. These constraints cut away pieces of the lattice to reveal a viable search space to run the minimization function. Finding the minimum point within that constrained lattice provides us with a matrix of the optimally matched jobs and sailors.

The constraints we apply are the same as the college admissions problem [5], where each job can fill up to but not exceed the allocated number of positions and each sailor can only be assigned to one position. The explicit program can be found in Appendix B.

4 Metrics

4.1 Pre-Match Metrics

4.1.1 Competitiveness

Definition

Competitiveness measures the relative desirability of a given sailor or job based on the expressed preference ranking of the other party. We wanted to consider a more sophisticated way of determining this metric beyond an average ranking for either the sailor or job as it is very susceptible to a right tail bias. For example, a position ranked 1st by ten individuals and hundredth by twenty individuals has the same average ranking as a job ranked seventh by all thirty individuals; yet the prior is much more competitive. To attempt to compensate for this, we adjust the average by a power of one half is to lessen the impact of very low preferences thereby weighting favorable preference more in our score consideration. In order to generalize the competitiveness score we scale the resulting average by the total number of jobs or sailors within the system. This allows for competitiveness to range from 1 being most competitive, and 0 being not competitive at all.

[math here]

Impact

Competitiveness will be used to rapidly identify the most attractive candidates. (E.g. Sailor A has a 0.98 competitiveness score and is therefore a top ranked candidate amongst all job owners.) These candidates would be good fits in many jobs and therefore could be considered ideal individuals to screen for command. On the job side, a low competitiveness would be a great way to decide on incentive structures. These jobs are those that very few sailors prefer and therefore either promotion based assignment or additional compensation could to targeted towards these positions.

4.1.2 Post-Match Metrics

Preference Allocation Windows Preference allocation indicates how many individuals received their top preference, how many individuals got a top three, top five, and top 10 preference, and how many failed to be matched with any preference.

Talent Distribution Some job owners have multiple jobs. Ideally, such job owners would get an approximately equal set of preferences; one owner with five job slots would not get their 1st-5th choices while another job owner with five job slots gets their 20th-25th preferences.

Preference Difference Gap This metric helps to illuminate the separation between job owner and sailor preferences. This can be an indicator for senior leaders that neither sailors nor job owners had their preferences unduly weighted within the optimization system.

5 Pilot

5.1 Overview

Overview of the pilot, still narrowing this in.

5.2 Data

5.3 Results

6 Alternate Formulations

6.1 Co-Location

Alternate to deterministic algorithms (such as Deferred Acceptance [2]) is a linear programming (optimization) approach. This allows system owners to input strategic objectives in the seeker-owner job matching process. The importance of job seeker preference can be weighted to be more important than owner, or vice versa. Requirements can also be added, as you will see in the formulation below.

The optimization function takes the form:

$$\begin{aligned}
\text{min} \quad & \sum_{i=1}^{i=n} \sum_{j=1}^{j=m} f(x_{i,j}) && (1)\\
\text{such that} \quad & \sum_{j=1}^{m} x_{i,j} \leq 1 \quad \forall i \in \{1, \dots n\} \quad \texttt{only one job per person} && (2)\\
& \sum_{i=1}^{i=n} \sum_{j=1}^{j=m} x_{i,j} == \min(n, m) \quad \forall j \in \{1, \dots m\} && (3)\\
& \quad \texttt{either all the jobs are filled or everyone has a job} && (4)\\
& \sum_{i=1}^{n} x_{i,j} \leq a_j \quad \forall j \in \{1, \dots m\} \quad \texttt{all jobs are at or below capacity} && (5)\\
& \frac{1}{n_c} \sum_{i=1}^{n_e} C(e_1, e_2) \geq 0.95 \quad \texttt{at least 95\% of couples are co-located} && (6)
\end{aligned}$$

The Goodness Function f is the strategic objective function of the assignment process. For the sake of this paper, we set it to value the preference of the seeker twice as much as the preference of th job owner.

$$f(x_{i,j}) = 2P^S_{i,j} + P^O_{j,i}$$

In matrix form this can be re-written (with $tr()$ indicating the trace):

$$
\begin{aligned}
\min \quad & tr(2X^TP^O) + tr(XP^S) & (7)\\
\text{such that} \quad & \sum_{j=1}^{m} x_{i,j} \leq 1 \quad \forall i \in \{1, \dots n\} \quad \texttt{only one job per person} & (8)\\
& X^T \bullet 1 \leq A \quad \texttt{all jobs are at or below capacity} & (9)\\
& \frac{1}{n_c}\sum_{i=1}^{n_e} X^TCXD \geq 0.95 \quad \texttt{at least 95\% of couples are co-located} & (10)
\end{aligned}
$$

Co-location Function C returns 1 if the couple is considered co-located, 0 if not or if single. Here we choose 50 miles between job locations to be consider co-located because that is the threshold for receiving dislocation allowance (DLA) for a permanent change of station (PCS) according to the Joint Travel Regulations (JTR). The location function $L(S_i)$ returns the lat/long location of the stationing for Seeker i.

$$
C(e_1, e_2) = \begin{cases} 0 & \text{if } e_2 == 0 \\ \mathbb{1}\big(||L(S_i) - L(S_j)|| \leq 50\big) & \text{otherwise} \end{cases}
$$

The inspiration and initial formulation of this optimization was done by a young Air Force officer who has since moved onto the private sector.

The extension of the formulation to include the Co-Location Function C and have more than one positions for each job a_i are the contributions of this team.

6.1.1 Explanation of Co-location constraint math

$$
\frac{1}{n_c}\sum\sum X^TCX \geq 0.95
$$

This works because

$$
\begin{aligned}
(CX)_{ij} &= \begin{cases} 1 & \text{Seeker } i\text{'s mate is assigned to job } j \\ 0 & \text{otherwise} \end{cases}\\
(X^T)_{ij} &= \begin{cases} 1 & \text{Seeker } i \text{ is assigned to job } j \\ 0 & \text{otherwise} \end{cases}\\
(X^TCX)_{ij} &= \begin{cases} n & \text{Number of couples assigned to the } (i,j) \text{ job pairing} \\ 0 & \text{otherwise} \end{cases}\\
D_{ij} &= \begin{cases} 1 & \text{The job pairing } (i,j) \text{ is considered co-location} \\ 0 & \text{otherwise} \end{cases}\\
(X^TCX)\dot{D} &= \text{ The Hadamard (element-wise) multiplication of } (X^TCX) \text{ and } D\\
\big((X^TCX)\dot{D}\big)_{ij} &= \begin{cases} n & \text{Number of co-located couples assigned to the } (i,j) \text{ job pairing} \\ 0 & \text{otherwise} \end{cases}\\
\sum\sum(X^TCX)_{ij} &= \text{Number of couples co-located}\\
\frac{1}{n_c}\sum\sum(X^TCX)_{ij} &= \text{Ratio of couples co-located}
\end{aligned}
$$

6.2 Tweaking Objective Function

Instead of a even weighting of job owner and seeker preferences in the objective function

$$f(x_{ij}) = x_{ij}(P_j^S + P_i^O)$$

Think of one where the Navy is seeking to enforce the "Needs of the Navy" upon the billeting process much more, weighting heavier the preferences of the job owners by a factor of ten

$$f(x_{ij}) = x_{ij}(P_j^S + 10P_i^O)$$

6.3 Wagering

Think instead of having the job seekers place preferences on all jobs, they have 100 points to distribute across all possible jobs. At the same time, job owners have 100 points for each opening

$$\begin{aligned} W^{S_i} &= \text{ Wagering vector of seeker } i \\ &\in [0, 100] \cap \mathbb{R}^{m\times 1} \\ \sum_j W_j^{S_i} &\leq 100 \\ f(x_{ij}) &= -x_{ij}(W_j^{S_i} + W_i^{O_j}) \end{aligned}$$

6.4 Wagering

6.5 Specialization Objective

7 Future Work

This is where our future work will go.

1. Explore implications of a "Separation" (u) preference
2. Explore strategic Importance of positions. This can either be an iteration where billets in priority tranche's are run iterative (tier 1 billets all matched, those sailors and billets are taken out of the pool, then tier 2 billets all matched, etc.). Or a weighting where in a single matching optimization the preferences of higher tiered billets are given a greater weighting
3. Does adding weight for specialization Im objective function help the Navy better? Did job owners who wanted specialized sailors get them? Did sailors who wanted specialized jobs get them?
4. Tailored Compensation descisions.
5. Incorporate timelines of expected roation date for availability windows.
6. Multiple firms and multiple workers, each sailor can opt into collaterals and other roles [4]

Appendices

A Notation

Throughout the paper, we will reference the notation listed in this section.

$$
\begin{aligned}
m &= \text{number of different jobs available} && (11)\\
n &= \text{number of persons} && (12)\\
\vec{P_i^S} &= \text{Preference vector of job seeker } i, \ \in \mathbb{Z}^{+,m\times 1} && (13)\\
P^S &= [\vec{P_i^S}|\dots|\vec{P_n^S}] \in \mathbb{Z}^{+,m\times n}, && (14)\\
& \quad \text{Preference Matrix of Seekers} && (15)\\
\vec{P_i^O} &= \text{Preference vector of job owner } j, \ \in \mathbb{Z}^{+,n\times 1} && (16)\\
P^O &= [\vec{P_i^O}|\dots|\vec{P_n^O}] \in \mathbb{Z}^{+,n\times m}, && (17)\\
& \quad \text{Preference Matrix of Job Owners} && (18)\\
\vec{A} &= \text{Position Available vector } \ \in \mathbb{Z}^{+,m\times 1} && (19)\\
a_j &= \text{Amount of positions for job } j, \ \in \mathbb{Z}^{+} && (20)\\
X &= \text{Placement Matrix} \in \{0,1\}^{n\times m} && (21)\\
x_{i,j} &= \begin{cases} 1 & \text{if } S_i \text{ is slated for job } j \\ 0 & \text{otherwise} \end{cases} && (22)\\
C &= \text{Co-Location Matrix, upper triangular} && (23)\\
C_{ij} &= \begin{cases} 1 & j > i, \ \text{and Seeker } i \text{ requests co-location with Seeker } j \\ 0 & j \le i, \ \text{or Seeker } e_{i,1} \text{ does not request co-location} \end{cases} && (24)\\
n_c &= \text{number of couples requesting co-location} && (25)\\
&= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij} && (26)\\
& && (27)
\end{aligned}
$$

B Matching Algorithms

Alternate to deterministic algorithms (such as Deferred Acceptance [2]) is a linear programming (optimization) approach. This allows system owners to input strategic objectives in the seeker-owner job matching process. The importance of job seeker preference can be weighted to be more important than owner, or vice versa. Requirements can also be added, as you will see in the formulation below.

The optimization function takes the form:

$$\min \quad \sum_{i=1}^{i=n} \sum_{j=1}^{j=m} f(x_{i,j}) \tag{28}$$

$$\text{such that} \quad \sum_{j=1}^{m} x_{i,j} \leq 1 \quad \forall i \in \{1, \dots n\} \tag{29}$$

$$\sum_{i=1}^{i=n} \sum_{j=1}^{j=m} x_{i,j} = \min(n, m) \quad \forall j \in \{1, \dots m\} \tag{30}$$

$$\sum_{i=1}^{n} x_{i,j} \leq a_j \quad \forall j \in \{1, \dots m\} \tag{31}$$

$$\tag{32}$$

The Goodness Function f is the strategic objective function of the assignment process. For the sake of this paper, we set it to value the preference of the seeker twice as much as the preference of th job owner.

$$f(x_{i,j}) = 2P^S_{i,j} + P^O_{j,i}$$

In matrix form this can be re-written (with $tr()$ indicating the trace):

$$\min \quad tr(2X^T P^O) + tr(XP^S) \tag{33}$$

$$\text{such that} \quad \sum_{j=1}^{m} x_{i,j} \leq 1 \quad \forall i \in \{1, \dots n\} \tag{34}$$

$$(X \bullet 1) \bullet 1 = \min(n, m) \quad \forall j \in \{1, \dots m\} \tag{35}$$

$$X^T \bullet 1 \leq A \tag{36}$$

$$\tag{37}$$

The constraints in lines (29) and (34) ensure that each seeker receives only one job, lines (30) and (35) ensure that either all the jobs are filled or everyone has a job, and lines (31) and (36) ensures that all jobs are at or below capacity (do not exceed capacity).

References

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

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

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

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

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

[6] S. Mitchell S. Khuller and V. Vazirani. On-line algorithms for weighted bipartite matching and stable marriages. *Theoretical Computer Science*, 127:255–267, May 1994.