

評価空間を用いた近似的な完全選好を推定するアルゴリズム
—大学における研究室配属問題への適用—

Algorithm for Estimating Perfect Preferences Utilizing an Assessment Area:
Application to Laboratory Assignment Problems

石上 諒* 岡田 勇† 篠宮 紀彦‡
Ryo Ishigami Isamu Okada Norihiko Shinomiya

1. はじめに

学生と学校、研修医と病院、男女の結婚といった 2 種類のエージェントがお互いの選好(好み)に基づいて最適な割当を行うマッチング問題は、経済学、人工知能、マルチエージェントシステム、特に選好に対しては心理学と幅広い分野で盛んに研究されている。理論的に検討されているマッチング問題は、Gale-Shapley アルゴリズム[1]を始め、両エージェントが持つ選好が完全に順序付けられている完全選好を前提にしている場合が多い。この完全選好という制約は、現実の多くの問題において達成が困難である。本稿で取り上げる学部における研究室配属問題でも、全ての学生や全ての教員に対する選好を表明する必要があり、厳密な順位付けは不可能である。不完全選好は、情報の不足、将来の不確実性、選択肢の複雑さ、時間あるいはリソースといった制約によってしばしば生じる。しかしながら、不完全選好を基にしたマッチングアルゴリズムの研究では、誤った順位や情報の欠如などを理由に局所的な最適解に陥り、結果としてマッチ全体が不安定になり、満足度が低くなるのが指摘されている[2-5]。ここで、マッチの満足度は、マッチした相手が選好リストの位置によって定まり、なるべく上位の相手とマッチした方が満足度は高い。また、マッチの安定性とは、マッチ相手に対して不満を持たないこと、参加者がペアを変えたいと思わないというパレート効率性を意味する。したがって、学生が選好リストに含めていない教員とマッチした場合、他に好ましい教員がいることを感じると、安定性と満足度が損なわれてしまう。以上から、従来の研究の共通の弱点は完全選好にできてないことに由来していると考えられる。未開拓の分野である選好の推定において、本研究では、学生の真の選好に近い完全選好を推定し、その下でマッチングを行い、結果的にマッチの安定性と全体の満足度を向上させることを目指す。

2. 提案アルゴリズム

研究室配属問題を例に、特徴ベクトルによって評価空間上に教員を配置し、学生にとって真の選好を表すベクトル量を推定することで、そこからの距離によって全ての教員の順位を求める。この手法により安定マッチングアルゴリズム(Gale-Shapley アルゴリズム)を適用することが可能となり、安定なマッチを作ることに貢献できる。選好の偏りや、様々な種類の研究室配属問題の制約によらずに不完全な選好から完全な選好を独自の手法で推定し、それを元に完全マッチングアルゴリズムに適用することで、満足度の高い結果を得られることをシミュレーションから確認した。

* 創価大学理工学部情報システム工学科 4 年

† 創価大学経営学部

‡ 創価大学理工学部

2.1 完全選好推定アルゴリズム

本研究が想定するモデルの概要を述べる。研究室数(=教員数) N に対して、 M 人の学生が配属希望を出す。各学生は上位 $k (< N)$ 位までの希望研究室リストを提示するものとする。もし $k = N$ ならば完全選好となるが、ここでは $k < N$ となる不完全選好を想定する。このときに、各学生の完全選好を推定するアルゴリズムを以下のように定義する。

手順 1. 教員を評価するために d 個の客観指標を定義し、全教員をその指標で測定し、 d 次元ユークリッド空間(これを評価空間と呼ぶ)上に配置する。教員 i の評価空間上の点を r^i とする。例えば、[論文数, 所属学生数, h-index]ならば $d = 3$ である。

手順 2. 学生ごとに次の動作を行う。

2-1. 希望研究室リストと整合する評価空間の点の集合 A_j を求める。すなわち希望リストが p 位の教員を n_p としたとき、 $A_j = \{x \mid \|r^{n_1} - x\| < \|r^{n_2} - x\| < \dots < \|r^{n_p} - x\|\}$ である。ここでノルム $\|x\|$ は $\|x\| = \max_i |x_i|$ である。ただし l とは d 次元ベクトルの成分番号($1 \leq l \leq d$)である。

2-2. A_j の中で第 1 希望の教員 n_1 に関する r^{n_1} に最も近い点を学生の選好点 a_j とする。

2-3. a_j から各教員の点 r^i の距離について、その距離の短い順に教員を並べたものを、学生の完全選好リストとする。

3. 推定性能の評価

提案アルゴリズムの有効性を実証するために、真の完全選好と推定した完全選好の比較をシミュレーションにより行った。評価空間にランダムに配置された教員の点を真の教員の点の位置、ランダムに配置された学生の選好点を真の選好点とした。真の教員の点から二次元ガウス分布に沿って移動させた教員の点と学生の真の選好点からの距離によって求められる順位を元に、学生が最初に提出する選好が決まる。そのサイズ k ごとに推定される推定完全選好と、設定した真の完全選好との乖離を、ケンドールの順位相関係数を用いて計算した。実際の創価大学の理工学部情報システム工学科の研究室の数である $N = 15$ を使用し、二次元ガウス分布の標準偏差 $\sigma = 0.02$ と設定して 1000 回シミュレーションを行った結果を下図 1 に示す。縦軸のケンドールの順位相関係数は $-1 \leq \tau \leq 1$ で定義され、1 に近づくほど、2 つの順位に相関があることを表す。#C は一致するペアの数、#D は一致しないペアの数である。更に、 σ の値を変動させたときの 2 つの推定手法による結果を下図 2 のように比較した。

$$\tau = \frac{(\#C - \#D)}{\frac{1}{2}n(n-1)}$$

Kendall's tau coefficient

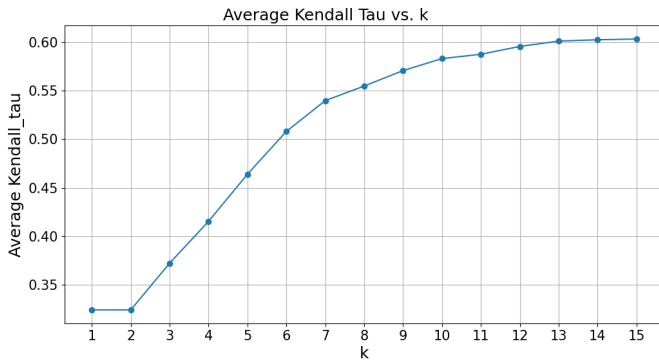
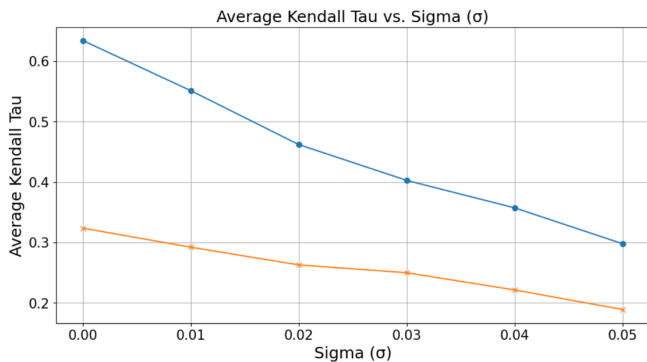


図 1 ケンドール順位相関係数による性能評価

図 2 σ における 2 つの推定手法の性能 ($k=5$)

4. 考察

図 1 から、学生が提供する教員の選好順位リスト k が増加するほど、学生の真の選好リストと推定される選好リストとの順位相関係数が増加することが分かる。ただし教員の客観値 r^i が学生の考える真の客観値から二次元ガウス分布 (σ) の量だけずれているため $N = k = 15$ でも 1 に到達しない。また、 k の値が増えても順位相関係数があまり上昇しないことから、適切な k の値で十分な推定が可能であることが示唆される。例えば図 1 によれば、 $k = 5$ の時で、 $= 15$ で得られる順位相関係数の 80% 程度の推定が可能となっている。

図 2 中、青線 (Estimation Method 1) とは、学生が自身の真の選好に対し 1 位から $k = 5$ 位の研究室 (教員) のリストを提示した場合を示し、黄線 (Estimation Method 2) とは、 N 人からランダムに選んだ k 研究室について、その 1 位から $k = 5$ 位までの順位について、自身の真の選好と整合するように並べた場合を示す。図 2 から、教員の真の評価値が見かけの評価値よりどのくらい乖離しているか (σ) の程度が小さいほど、順位相関係数が改善することが分かる。また、Estimation method 1 の方が 2 よりも順位相関係数を 1.5-2 倍程度改善させられることから、学生が不完全選好リストを提示する場合、その学生にとって、なるべく上位に評価される教員を優先的に提示するように仕向けることが重要であることも分かる。

5. おわりに

本論文では、不完全な選好が存在する研究室配属問題において完全マッチングアルゴリズムを適用するために、不完全選好から完全選好を推定する手法を提案した。その際に現実的な時間スケールで実行可能となるような推定アルゴリズムを設計した。シミュレーションを用いてアルゴリズムの推定性能を評価し、エージェントのサイズに対し、ある程度少ないサイズの選好を提出するだけで、潜在的な真の選好に近い完全選好を予測することができた。これにより、推定した完全選好の下でマッチングを行うことで、不完全選好下でマッチングを行った場合よりもマッチ全体の満足度が上昇することは明らかだと分かる。本研究の成果は、経済学と計算機科学、そして心理学を横断する手法であるといえるため、様々な分野への重要なインパクトを持つと考えられる。特に、評価手法に関して、シミュレーションによるアルゴリズムの推定性能評価だけでなく、学生にこのアルゴリズムを使用した web サイトを使用してもらい、実際の学生データに基づいた選好推定、満足度評価を行えることも、この研究の優れた側面である。

将来、研究室配属を運営する組織側と選好を提出する学生側のどちらにとっても、アルゴリズムによって両者の負担が軽減されることが期待される。そのためにも、これから注力すべき本研究の課題についていくつか触れておく。第 1 に、サンプルサイズが比較的小さく、結果の一般化に限界がある。第 2 に、本研究で使用した分析手法にはいくつかの仮定が含まれており、それらの仮定が結果に影響を与える可能性がある。したがって今後の研究では、まず、サンプルサイズを増やし、より多様なケースからデータを収集することで、結果の一般化可能性を高める。その際、大規模データセットにも対応できるような効率的なアルゴリズム設計を目指す。次に、現在の分析手法の限界を克服するために、異なる統計手法を検討したり、実際に実験を実施したりを通して、結果の頑健性を検証する。これらにより、研究の精度と信頼性を向上させるとともに、より広範な適用可能性を検討できることが期待される。

参考文献

- [1] D. Gale, L.S. Shapley, "College admissions and the stability of marriage," *The American Mathematical Monthly*, vol.69, no.1, pp.9-15, Jan. 1962.
- [2] 武田 雅俊, 鈴木 伸崇, "選好の有無が混在する場合のマッチング手法," *DEIM Forum 2023*, no.1, pp.1-2, June. 2023.
- [3] 田邊 優斗, 鈴木 伸崇, "二部グラフにおける選好の拡張方法について," *DEIM Forum 2023*, no.1, pp.2-8, June. 2023.
- [4] 高根澤 武蔵, 真鍋 義文, 多対二の研究室配属問題," *情報処理学会第 84 回全国大会* pp.1, 2023.
- [5] Rastegari, B., Condon, A., Immorlica, N., & Leyton-Brown, K, "Two-sided matching with partial information," pp.733-750, June. 2013.