

SIMILARITY-BASED ALLOCATION OF ACADEMIC SUPERVISORS UNDER PREFERENCE UNCERTAINTY

Martin Ndithi Ndeto

1. Problem Statement

Allocating students to academic supervisors is a recurring coordination problem in higher education institutions. The process typically relies on manual judgment informed by limited preference information, which can result in imbalanced supervisory workloads, inconsistent assignment criteria, and mismatches between student research interests and supervisory expertise.

From a research perspective, this setting exhibits several characteristics common to institutional allocation problems: incomplete preference information, capacity constraints, and the absence of an objectively defined optimal outcome. These characteristics limit the applicability of purely optimization-based approaches and motivate the use of interpretable, decision-support models.

This memo investigates the following research question:

Can similarity-based models support student-supervisor allocation under preference uncertainty and capacity constraints while improving consistency and perceived match quality?

The intent is to study the behavior and limitations of simple, interpretable allocation mechanisms in institutional settings rather than to automate academic decision-making.

2. Data Description

2.1 Student Profiles

Student data consisted of structured profiles collected at the point of allocation. Attributes included:

- Declared research interests and thematic keywords
- Academic background and prior coursework
- Project or topic preferences

The data exhibited substantial heterogeneity: some students provided detailed research descriptions, while others supplied only coarse or generic preferences.

2.2 Supervisor Profiles

Supervisor profiles included:

- Declared research areas
- Departmental affiliation
- Supervision capacity for the allocation cycle

Compared to student profiles, supervisor data was relatively stable and lower dimensional.

2.3 Data Characteristics and Constraints

The dataset exhibited several limitations:

- No ground-truth “optimal” assignment existed
- Preferences were self-reported and noisy
- Satisfaction outcomes were not directly observed

These constraints necessitated evaluation through proxy metrics and comparative analysis rather than accuracy against labeled outcomes.

3. Methodology

3.1 Feature Representation

Student and supervisor profiles were embedded in a shared feature space using normalized categorical attributes and keyword-based representations. This representation allowed heterogeneous profile information to be compared using standard distance measures while preserving interpretability.

3.2 Similarity-Based Matching

Similarity between students and supervisors was computed using distance-based measures over the shared feature space. A k-Nearest Neighbors (KNN) approach was used to identify supervisors whose profiles most closely aligned with a given student's interests.

KNN was selected due to its transparency, minimal modeling assumptions, and suitability for small to medium-scale institutional datasets, consistent with prior work on similarity-based matching and recommender systems.¹

3.3 Capacity-Constrained Assignment

Raw similarity rankings were post-processed to enforce supervisory capacity constraints. When a supervisor reached capacity, subsequent assignments were redirected to the next most similar available supervisor. This decoupling of similarity estimation from constraint enforcement enabled inspection of trade-offs between match quality and load balance.

4. Evaluation

4.1 Evaluation Strategy

In the absence of labeled optimal matches, evaluation emphasized structural and behavioral indicators rather than absolute correctness. The focus was on understanding how the allocation mechanism behaved under realistic institutional constraints.

4.2 Metrics

Evaluation considered:

- Variance in supervisory workload
- Frequency of post-allocation manual reassignment
- Qualitative student feedback on perceived alignment

These metrics served as proxies for allocation quality and institutional acceptability.

4.3 Findings

Similarity-based allocation reduced workload imbalance relative to heuristic and manual baselines. Assignments generated through similarity scoring required fewer administrative adjustments and were perceived as more aligned with student research interests, particularly in cases where preference information was moderately detailed.

5. Discussion

5.1 Interpretation

The results suggest that similarity-based methods can function effectively as decision-support tools in institutional allocation settings. The separation of similarity estimation from constraint handling enabled transparent inspection of allocation trade-offs and facilitated human oversight.

5.2 Limitations

The approach does not guarantee global optimality and does not explicitly encode fairness objectives, such as balancing seniority or demographic considerations. Additionally, reliance on self-reported preferences introduces bias and uncertainty into similarity estimates.

5.3 Research Directions

Future research directions include:

- Formulating the problem as a constrained bipartite matching task
- Incorporating fairness-aware objectives into the allocation process²
- Studying longitudinal satisfaction and outcome measures

6. Concluding Note

This memo illustrates how simple similarity-based models can serve as research instruments for studying allocation behavior under preference uncertainty and capacity constraints. The work highlights opportunities for principled investigation of institutional matching problems using interpretable methods.

References

1. Ricci, F., Rokach, L., & Shapira, B. (2011). *Introduction to recommender systems handbook*. Springer.
2. Abdulkadiroğlu, A., & Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*.

Status Note

This memo documents exploratory research derived from an operational academic allocation process and is shared to provide context on problem formulation, methodology, and evaluation under real-world institutional constraints. The work has not yet been submitted for peer review and continues to evolve.