STUDIA INFORMATICA
Nr 1 (30) Systems and information technology 2024

Aleksandra Zakrzewska[1]

University of Lodz
Faculty of Mathematics and Computer Science
ul. Banacha 22, 90-238 Łódź, Poland

[1]aleksandra.zakrzewska@wmii.uni.lodz.pl

A modification of the Gale-Shapley algorithm used to assign Polish pupils to schools

DOI: 10.34739/si.2024.30.07

Abstract. Allocation algorithms based on rankings are utilized in various situations, such as selecting appropriate tasks for workers, scheduling doctors' appointments, or assigning children to schools. A modified version of the standard Gale-Shapley algorithm is introduced and employed in Poland's secondary school recruitment process. This algorithm relies on students' school results and their priority lists. During the recruitment process, students are assigned to a school on the basis of their performance and preference list. However, this approach disadvantages weaker students, as they are only considered towards the end of the process. Consequently, they are often assigned to schools with a lower priority on their lists or, in some cases, remain unassigned. The proposed modification aims to improve opportunities for students to plan their educational future more fairly, considering their individual achievements.

Keywords: Allocation algorithm, Recruitment process, optimization, Gale-Shapley algorithm, Recruitment simulation

1 Introduction

1.1 Recruitment algorithms

The history of the recruitment algorithms is very long. Initially, it relied on selection algorithms, a well-known problem in computer science [11]. Suppose that there is a given data with assigned numerical values to each object, then identifying the top k elements can be easily achieved through sorting algorithms.

This problem finds applications in various domains, including hiring [6], [7], crowdsensing networks [8], [9] and even matchmaking for marriages [10].

Since the recruitment is typically conducted by algorithms rather than humans, one might assume that they would be more objective. However, this assumption has been proven false, and numerous authors have addressed this issue in their research [3], [4], [5].

1.2 School recruitment process

One application of recruitment algorithms is the admission of new students to schools, where schools typically aim to attract the most promising candidates.

In Poland during the 1990s [1] the recruitment process for secondary schools was school-dependent. Each school defined its own recruitment rules. This system has two major drawbacks. Firstly, the process lacked synchronization, allowing the same student to be admitted to multiple schools. As a result, if a student withdrew their acceptance, the entire recruitment process had to be repeated in that school. Secondly, the rules varied across schools, with some requiring exams while others conducted only interviews. Despite these inconsistencies, the system was relatively easy to implement at the school level, often involving manual sorting of student applications and subsequent announcement of assigned lists.

In present-day Poland, the recruitment process has been synchronized within administrative areas such as cities, with standardized rules across all schools [2]. These rules now depend on the students' results. To achieve an optimal outcome, the Gale-Shapley algorithm [10] is employed in the recruitment process. The algorithm used in Poland has also been analyzed by [12] and [13]. The theoretical explanations of the official algorithm can be found in these works.

In Poland, teenagers typically complete primary school around the age of 14. In their final year (8th grade), they are required to take external exams in Polish, Mathematics, and a foreign language. The exam results, along with their grades from the primary school and additional achievements, are converted into a points system ranging from 0 to 200 [2]. Preferred secondary schools are chosen based on calculated points. The point calculation depends on the class they have chosen, so the final score can vary for each student across different classes. This is the first drawback of the recruitment process: after selecting a school, calculating the points, and comparing them with historical data, the process must be repeated for each class. Furthermore, students must prepare a priority list of schools before even taking the external exams, requiring them to predict their own results. This fair but complex system poses a challenge for young people to comprehend. This issue is particularly evident among weaker students who prioritize higher-ranking schools on their list but ultimately fail to get anywhere.

1.3 Points in the recruitment system

The recruitment process outcome is determined by two factors: school results (100 points) and external exam results (100 points).

The results from the external exams are provided in percentages, denoted by r, which are then converted to points using the following formulas:

- Mathematics and Polish: - $0.35 * r$,
- A foreign language - $0.3 * r$.

Each secondary school class creates a list of four subjects that are considered in the process. Two of these subjects are always Mathematics and Polish, while the remaining two can be chosen freely from all subjects taught at the school. It can be a single subject like Geography or a combination of several subjects like Biology/Geography/History. The grades (ranging from 2 to 6) obtained in each of these four subjects are converted into points. If there is a combination of subjects instead of a single subject, the algorithm selects the highest grade. The conversion of grades to points is as follows: $6 \rightarrow 18$ points, $5 \rightarrow 17$ points, $4 \rightarrow 14$ points, $3 \rightarrow 8$ points, and $2 \rightarrow 2$ points.

Furthermore, students can earn additional points: 7 points for achieving a certificate with distinction (if the average grade is 4.75 or higher), 3 points for volunteering, and 18 points for additional achievements such as competition participation.

To facilitate the school selection process, schools share historical recruitment data with students. For each class, information regarding the minimum, maximum, and average number of points obtained by students assigned to that class in the previous year is provided. This data enables the creation of a school ranking, aiding students in their decision-making process.

1.4 Recruitment algorithm

The recruitment algorithm used in the admission process is presented in Algorithm 1. The algorithm considers students one by one and attempts to assign each student to their top-preferred class. If the class has available places, the student is assigned to it. If the class is full, a comparison is made between the student's results and the results of the last student already assigned to that class. If the student's results are worse, the algorithm moves on to the next class in the preference list. If the student's results are better, he/she is assigned to the class, and the previously assigned student is returned to the pool of students to be considered. Each class is evaluated separately, as they have their own subject lists. In case of a tie, additional rules are considered, such as a student's disability or being from a large family.

Sorting the list of students by their results enhances the effectiveness of the algorithm. This change helps minimize the need for reassignment (returning the last student to the list of students to be considered).

During the recruitment process, the maximum size of the preferred list must be specified. In Lodz, for example, in 2021 and previous years, students were limited to choose three classes. However, starting from 2022, this limitation was removed, and students are now allowed to choose any number of classes.

Algorithm 1 Standard recruitment algorithm

Require: List of all classes *classes* with the 4 subjects
Require: List of all students *students* with their preference list (of classes)
Ensure: all students are assigned to one of the classes or are in the *unassigned* set

```
while students is not empty do
    student is a first element of students
    for class in student.preferedList do
        if class is not full then
            assign student to class
            remove student from students
        else
            last is the student already assigned to class with the lowest number of points (calculated for class)
            if student has more points than last (in case of a tie choose the winner using additional rules) then
                assign student to class
                remove student from students
                add last to students
            end if
        end if
    end for
    if student is not assigned to any class then
        add student to unassigned
        remove student form students
    end if
end while
```

2 The modification of the recruitment algorithm

2.1 Motivation

This section presents a modification of the recruitment algorithm aimed to address some of the limitations observed in the current system. By adding an additional rerecruitment process, the proposed modification provides more balanced opportunities for students, especially those with lower scores, to secure placements in schools that align better with their preferences. The following subsections detail the motivation behind the modification and the adjustments made to the algorithm's structure.

While a good student should have no difficulty in getting into their chosen school, it becomes challenging to predict where a student with lower results is assigned. Such students often end up unassigned (if they selected top schools) or assigned to weaker schools (chosen as a last place in their priority list). Consequently, after the recruitment process, the high-performing students are content with their outcomes, while the rest may be dissatisfied.

Following the initial recruitment, a standard rerecruitment process takes place. Each school has a list of available places and can recruit students to fill those vacancies. Occasionally, even in prestigious schools, students may withdraw, creating an opportunity for the school to admit another person. The specifics of this rerecruitment process vary among schools, but it typically follows a FIFO rule.

In this article, an additional recruitment phase is introduced. It should take place immediately after the initial recruitment and prior to schools initiating their rerecruitment processes. This supplementary recruitment process is primarily based on the extended subject offerings in secondary schools.

2.2 Teaching in the secondary school in Poland

In Poland, upon completing secondary school, every student is required to take the matura exam. This exam comprises three mandatory subjects at the basic level: Polish, Mathematics, and a foreign language. Additionally, each student must choose from one to six subjects at the extended level, with the choice typically based on their intended future education.

In secondary school classes, most subjects are taught at the basic level, except for a few that are taught at the extended level. This distinction is evident even in the first year of secondary school. For example, a basic-level Biology class may have one lesson per week, whereas an extended-level Biology class could have three lessons per week (though the specific hours may vary depending on the school). Consequently, when students create their preference lists for class selection, they also consider which subjects are offered at the extended level.

2.3 The algorithm

In this approach, each student, while creating their preference list, also selects two subjects they wish to study at the extended level. See the Algorithm 2. If information about previous

Algorithm 2 Rerecruitment algorithm

Require: List of all classes *classes* (not full) with information about extended subjects
Require: List of all unassigned students *students* with 2 chosen subjects
Ensure: all students are assigned to one of the classes or are in the set of unassigned *unassigned*

```
sort students by their results
while students is not empty do
    student is a first element of students
    for class in classes do
        if class is not full then
            if class has at least 1 subject chosen by the student then
                assign student to class
                remove student from students
            end if
        end if
    end for
    if student is not assigned to any class then
        add student to unassigned
        remove student form students
    end if
end while
```

year results is available, all classes can be further sorted in a manner that places higher-ranked schools at the top. This enables the commencement of the rerecruitment process with better students and better schools.

3 Empirical result

The performance of the algorithm is evaluated using a Monte Carlo simulation approach, employing two sets of data: random data and real data obtained from the Lodz City Hall (Urzad Miasta Łodzi) from 2022.

For the random data, a Monte Carlo simulation is conducted by generating 1000 students, 25 schools, and 2 classes per school, with each class limited to 20 students. Student results are assigned randomly with equal probability.

For the real data, information was obtained from the Lodz City Hall, covering 6876 students, 56 schools, and a total of 290 classes with varying class limits of 15, 20, and 30 students. Given the capacity, there were enough places to accommodate all students. Further, the average number of points obtained by accepted students in 2021 is assigned to each school. This allows a Monte Carlo analysis of algorithm performance under real-world conditions.

3.1 Monte Carlo Simulation 1: Impact of Preference List Size

The objective of this Monte Carlo simulation is to examine how the size of the preference list affects the recruitment process. Monte Carlo methods are used to conduct multiple trials, adding variability and robustness to the results by randomizing aspects of the recruitment process. Specifically, the size of the preference list is varied from 1 to 10 across multiple iterations. Additionally, one simulation is performed without any limits on the preference list, with the size randomly determined for each student. The percentage of unassigned students is then analyzed. Each simulation is repeated 10 times for the random data set.

For the real data, the simulation was repeated twice. In the first run, classes are randomly assigned from the entire list for each student. In the second run, classes were assigned based only on schools whose average points are within 30 points of the student's result, mimicking a real-world scenario where high-performing students avoid weaker schools. This scenario is called "real data with modification".

The results are presented in Table 1. Based on the simulations, the following conclusions

Table 1. The percentage of unassigned students. The size is the size of the preference list. Real data refers to the situation that classes are random. In real data with modification the classes are assigned based on results.

size	all data	random data	real data	real data with modification
1	13%	3%	10%	21%
2	9%	1%	7%	19%
3	7%	1%	6%	19%
4	7%	1%	6%	19%
5	7%	1%	6%	18%
6	7%	0%	6%	19%
7	7%	0%	6%	19%
8	6%	0%	6%	18%
9	6%	0%	6%	18%
10	6%	0%	6%	18%
no limit	6%	0%	6%	18%

can be drawn:

- A preference list size of 3 classes appears optimal. Increasing the list size or removing limits has minimal effect on outcomes.
- Randomized simulations without context lack credibility and may not represent real-world dynamics accurately.
- Limiting students' class choices (real data with modification) can produce significantly different outcomes than a simulation, where classes are chosen randomly. Without this additional rule it may happen that some very good students may end up in schools with weaker reputations. The simulation with this modification is much more credible.

For the subsequent simulation, only the case where the number of classes in the preference list is limited to 3 is considered.

3.2 Monte Carlo Simulation 2: Effectiveness of Modified Recruitment Process with Rerecruitment

In this simulation, the effectiveness of the modified recruitment process with the inclusion of rerecruitment is examined. The percentage of unassigned students with and without rerecruitment is compared. The simulation is conducted 10 times, and the results are presented in Table 2 The results indicate that rerecruitment assists with allocating the students. When

Table 2. The percentage of unassigned students with and without rerecruitment.

data	without rerecruitment	with rerecruitment
random data	0.85%	0.61%
real data	1.06%	0.36%
real data with modification	19.6%	17.92%

considering real data with modification, where students are assigned to classes based on their results, approximately 18% of students remain unassigned. However, with the rerecruitment process, around 2% of students (115 people) are additionally able to secure their preferred schools.

4 Conclusion

In this study, the recruitment process for secondary schools in Poland is analyzed. The existing recruitment algorithm is highly objective, as it relies solely on students' result, minimizing subjective factors in the allocation process. The algorithm performs well with random data and even when schools/classes are randomly chosen. However, in real-world scenarios where students choose their classes based on their results, approximately 19% of students are left unassigned. Although the presented rerecruitment process improves the outcome by approximately 2 percentage points, this adjustment alone is not sufficient to ensure that all students are assigned to their preferred schools.

The main challenge lies in the limited number of places available in schools. Even if the total number of places exceeds the number of students, students tend to choose better schools, resulting in many classes in weaker schools remaining unfilled due to low enrollment. This situation may change in the future with a decrease in the number of students, as Poland is currently experiencing a low birth rate.

Additionally, the circumstances presented in this study are also unique, as the recruitment process in 2022 and 2023 in Poland includes children born not only within a single year but within one and a half years. As a result, there is approximately 1.5 times more students than usual, while the number of students that can be admitted to schools is limited by schools' size. However, analyzing the problem using such data, where there are 7570 places available in schools for 6876 students (1.1 times more), provides a better understanding of the issue compared to scenarios where there are twice as many places as students.

In the future, enhancements to the recruitment algorithm could include advanced techniques, such as neural networks or machine learning models, which could further optimize the process. Such methods may help in identifying patterns or preferences that could improve fairness and efficiency, potentially offering students even better chances of admission to their preferred schools while maintaining the objectivity and transparency of the current system.

References

1. Dz.U. Rozporzadzenie Ministra Edukacji Narodowej z dnia 9 maja 1992 r. w sprawie warunków przyjmowania uczniów do szkół publicznych oraz przechodzenia z jednych typów szkół do innych. Dziennik Ustaw (Polish Journal Of Laws). **42** (1992)
2. Dz.U. Rozporzadzenie Ministra Edukacji Narodowej z dnia 18 listopada 2022 r. w sprawie przeprowadzania postepowania rekrutacyjnego oraz postepowania uzupełniajacego do publicznych przedszkoli, szkół, placówek i centrów. Dziennik Ustaw (Polish Journal Of Laws). **2431** (2022)
3. Köchling, A. & Wehner, M. Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. Business Research. **13**, 795-848 (2020)
4. Solon, B. & D., S. Big Data's Disparate Impact. 104 California Law Review. **104:671** pp. 671-732 (2016)
5. Lepri, B., Oliver, N., Letouzé, E., Pentland, A. & Vinck, P. Fair, Transparent, and Accountable Algorithmic Decision-making Processes. Philosophy & Technology. **31**, 611-627 (2018)
6. Asidik, I., Kusrini & Henderi Decision Support System Model of Teacher Recruitment Using Algorithm C4.5 and Fuzzy Tahani. Journal Of Physics: Conference Series. **1140**, 012030 (2018)
7. Faliagka, E., Ramantas, K., Tsakalidis, A. & Tzimas, G. Application of Machine Learning Algorithms to an online Recruitment System. (2012,1)
8. Campioni, F., Choudhury, S., Salomaa, K. & Akl, S. Improved Recruitment Algorithms for Vehicular Crowdsensing Networks. IEEE Transactions On Vehicular Technology. **68**, 1198-1207 (2019)
9. Yi, K., Du, R., Liu, L., Chen, Q. & Gao, K. Fast participant recruitment algorithm for large-scale Vehicle-based Mobile Crowd Sensing. Pervasive And Mobile Computing. **38** pp. 188-199 (2017)
10. Gale, D. & Shapley, L. College Admissions and the Stability of Marriage. The American Mathematical Monthly. **69**, 9-15 (1962), http://www.jstor.org/stable/2312726
11. Cormen, T., Leiserson, C., Rivest, R. & Stein, C. Introduction to Algorithms, 3rd Edition. (MIT Press,2009)
12. Świtalski, Z. Optymalny system rekrutacji kandydatów do szkół. Badania Operacyjne I Decyzje. **3–4** pp. 85-98 (2006)
13. Urbański, A. & Nawrocki, J. Algorytmy sprawiedliwej kwalifikacji uczniów i studentów w szkołach średnich i wyższych. Acta Universitatis Lodziensis Folia Oeconomica. **167**, 413-425 (2003)