IE-400 Principles of Engineering Management

Group 11

Project Report

Job Matching Optimization:
Maximizing Priority and Minimizing Dissimilarity

Group Members:

Emre Furkan Akyol	22103352
İbrahim Çaycı	22103515
Batu Uzun	22103810

11.05.2025

Contents

1 Introduction

This report details the methodology and results for a two-part job matching optimization problem. The primary goal is to effectively match job seekers with available job openings. Part 1 of the problem focuses on maximizing the total priority weight of filled positions, considering various compatibility criteria between seekers and jobs. Part 2 builds upon this by aiming to minimize the maximum dissimilarity score among the assignments, while still achieving a significant portion (defined by a parameter ω) of the maximum priority weight found in Part 1. Integer Linear Programming (ILP) models were developed and implemented using Gurobi to solve both parts of this problem.

2 Methodology

2.1 Data Preprocessing

Before formulating the ILP models, several preprocessing steps were performed on the ‘seekers.csv‘ and ‘jobs.csv‘ datasets:

- **List Parsing:** Columns containing lists (e.g., "Skills", "Required_Skills", "Questionnaire" for both seekers and jobs) were parsed from their string representations into actual Python list objects using ‘ast.literal_eval‘.
- **Experience Level Mapping:** Text-based experience levels (e.g., "Entry-level", "Mid-level") were mapped to numerical values (1 to 5) for both seekers and jobs to facilitate comparison. For example: "Entry-level" $\rightarrow$ 1, "Mid-level" $\rightarrow$ 2, "Senior" $\rightarrow$ 3, "Lead" $\rightarrow$ 4, "Manager" $\rightarrow$ 5. These were stored in ‘Experience_Level_Mapped‘ and ‘Required_Experience_Level_Mapped‘ columns.
- **Location Distances:** The ‘location_distances.csv‘ file was loaded to provide commute distances between different locations.
- **Job Type Mapping:** All unique job types from seekers' desired types and jobs' actual types were collected and mapped to numerical indices (0 to $N_{types} - 1$) to facilitate comparisons.
- **Skill Vectors:** A comprehensive list of all unique skills was created. For each seeker i and job j, binary mappings ($m_{i,sk}$ and $n_{j,sk}$) were created indicating possession or requirement of each skill sk.

2.2 ILP Model - Part 1: Maximize Priority-Weighted Matches

The objective of Part 1 is to find a set of seeker-job assignments that maximizes the sum of priority weights of the filled job positions, subject to compatibility and resource constraints.

2.2.1 Indices and Sets

- S: Set of job seekers, indexed by i.

- J: Set of job openings, indexed by j.
- $AllSkills$: Set of all unique skills, indexed by sk.

2.2.2 Parameters

- W_j: Priority weight of job j.
- P_j: Number of available positions for job j.
- $JobTypeIndex_{s,i}$: Numerical index of the desired job type of seeker i.
- $JobTypeIndex_{job,j}$: Numerical index of the type of job j.
- $MinDesiredSalary_i$: Minimum desired salary of seeker i.
- $SalaryRangeMin_j, SalaryRangeMax_j$: Salary range (minimum and maximum) for job j.
- $m_{i,sk}$: Binary, 1 if seeker i has skill sk, 0 otherwise.
- $n_{j,sk}$: Binary, 1 if job j requires skill sk, 0 otherwise.
- $ExpLevel_i$: Experience level (mapped numerically) of seeker i.
- $ReqExpLevel_j$: Required experience level (mapped numerically) for job j.
- $IsRemote_j$: Binary indicator, 1 if job j is remote, 0 otherwise.
- $Location_i$: Location of seeker i.
- $Location_j$: Location of job j.
- $MaxCommute_i$: Maximum commute distance for seeker i.
- $Dist(loc_s, loc_j)$: Distance between seeker i's location and job j's location.
- M_{TYPE}: A Big-M constant for job type compatibility. It is set to the maximum possible absolute difference between job type indices (i.e., the total number of unique job types minus one). This value ensures that if $x_{ij} = 0$, the job type constraints are non-restrictive, and if $x_{ij} = 1$, they enforce equality of job type indices.
- M_{SALARY}: A Big-M constant for salary compatibility. It is chosen as the maximum value found across all seekers' minimum desired salaries ('Min_Desired_Salary') and jobs' maximum salary range ('Salary_Range_Max'). This ensures it is large enough to make the salary constraints non-restrictive when $x_{ij} = 0$.
- M_{EXP}: A Big-M constant for experience compatibility. This is set to 5. Since the experience levels are mapped from 1 to 5, the maximum possible absolute difference between a seeker's experience and a job's required experience is 4. A value of 5 for M_{EXP} ensures the constraint $ExpDiff_{ij} + M_{EXP}(1 - x_{ij}) \geq 0$ correctly enforces that the seeker's experience level is greater than or equal to the job's required level ($ExpLevel_i \geq ReqExpLevel_j$) when an assignment is made ($x_{ij} = 1$).

- M_{DIST}: A Big-M constant for location compatibility. This is set to the maximum distance found in the location distances matrix. This value is used to make the location constraint non-restrictive when $x_{ij} = 0$, or when $x_{ij} = 1$ and the job is remote.

2.2.3 Decision Variables

- $x_{ij} \in \{0, 1\}$: Binary variable, 1 if seeker i is assigned to job j, 0 otherwise.

2.2.4 Objective Function

Maximize the total sum of priority weights of filled positions:

$$\text{Maximize } Z = \sum_{i \in S} \sum_{j \in J} W_j \cdot x_{ij} \tag{1}$$

2.2.5 Constraints

An assignment x_{ij} can only be 1 if all compatibility conditions are met. These are enforced using Big-M formulations:

1. **Job Type Compatibility:** If $x_{ij} = 1$, then seeker i's desired job type must match job j's type.

$$JobTypeIndex_{s,i} - JobTypeIndex_{job,j} \leq M_{TYPE}(1 - x_{ij}) \qquad \forall i \in S, j \in J \tag{2}$$
$$JobTypeIndex_{job,j} - JobTypeIndex_{s,i} \leq M_{TYPE}(1 - x_{ij}) \qquad \forall i \in S, j \in J \tag{3}$$

2. **Salary Compatibility:** If $x_{ij} = 1$, seeker i's minimum desired salary must be within job j's range.

$$MinDesiredSalary_i - SalaryRangeMin_j + M_{SALARY}(1 - x_{ij}) \geq 0 \quad \forall i \in S, j \in J \tag{4}$$
$$SalaryRangeMax_j - MinDesiredSalary_i + M_{SALARY}(1 - x_{ij}) \geq 0 \quad \forall i \in S, j \in J \tag{5}$$

3. **Skills Compatibility:** If $x_{ij} = 1$, seeker i must possess all skills required by job j. For every skill $sk \in AllSkills$:

$$x_{ij} \leq m_{i,sk} + (1 - n_{j,sk}) \quad \forall i \in S, j \in J, sk \in AllSkills \tag{6}$$

(If job j requires skill sk (i.e., $n_{j,sk} = 1$), then $x_{ij} \leq m_{i,sk}$, forcing $m_{i,sk} = 1$ if $x_{ij} = 1$.)

4. **Experience Compatibility:** If $x_{ij} = 1$, seeker i's experience level must meet or exceed job j's required level. Let $ExpDiff_{ij} = ExpLevel_i - ReqExpLevel_j$.

$$ExpDiff_{ij} + M_{EXP}(1 - x_{ij}) \geq 0 \quad \forall i \in S, j \in J \tag{7}$$

5. **Location Compatibility:** If $x_{ij} = 1$, the job location must be acceptable to the seeker.

$$Dist(Location_i, Location_j) - MaxCommute_i \leq M_{DIST}(1 - x_{ij} + IsRemote_j) \quad \forall i \in S, j \in J \tag{8}$$

(If $IsRemote_j = 1$, the term $M_{DIST} \cdot IsRemote_j$ effectively makes the constraint non-binding for $x_{ij} = 1$. If $IsRemote_j = 0$ and $x_{ij} = 1$, it forces $Dist(Location_i, Location_j) \leq MaxCommute_i$.)

6. **Seeker Assignment Limit:** Each seeker can be assigned to at most one job.

$$\sum_{j \in J} x_{ij} \leq 1 \quad \forall i \in S \tag{9}$$

7. **Job Capacity Limit:** The number of seekers assigned to a job cannot exceed its available positions.

$$\sum_{i \in S} x_{ij} \leq P_j \quad \forall j \in J \tag{10}$$

2.3 ILP Model - Part 2: Minimize Maximum Dissimilarity

The objective of Part 2 is to minimize the maximum dissimilarity score among all assignments, while ensuring the total priority weight of assignments is at least $\omega\%$ of the maximum total priority weight (M_W) obtained from Part 1.

2.3.1 Indices and Sets

Same as Part 1.

2.3.2 Parameters

In addition to parameters from Part 1:

- M_W: Maximum total priority weight achieved in Part 1.
- ω: A percentage (e.g., 70, 80, 90) representing the target for total priority weight relative to M_W.
- d_{ij}: Pre-calculated dissimilarity score between seeker i's questionnaire and job j's questionnaire. $d_{ij} = \frac{1}{20} \sum_{k=1}^{20} |Quest_{seeker,i,k} - Quest_{job,j,k}|$.

2.3.3 Decision Variables

In addition to x_{ij} from Part 1:

- $M \geq 0$: A continuous variable representing the maximum dissimilarity score among all made assignments. (Range 0.0 to 5.0)

2.3.4 Objective Function

Minimize the maximum dissimilarity:

$$\text{Minimize } M \tag{11}$$

2.3.5 Constraints

Includes all compatibility and limit constraints (1-7 from Part 1 formulation above), plus the following:

8. **Minimum Priority Target:** The sum of priority weights of assigned jobs must be at least $\omega\%$ of M_W.
$$\sum_{i\in S}\sum_{j\in J} W_j \cdot x_{ij} \geq \frac{\omega}{100} \cdot M_W \tag{12}$$

9. **Maximum Dissimilarity Link:** If an assignment x_{ij} is made (i.e., $x_{ij} = 1$), its dissimilarity score d_{ij} must be less than or equal to M.
$$d_{ij} \cdot x_{ij} \leq M \quad \forall i \in S, j \in J \tag{13}$$

2.4 Implementation of Requirement Checks and Location Rule

The compatibility requirements (job type, salary, skills, experience, location) were implemented using the Big-M formulations described in constraints 1-5 of the Part 1 model. This embeds the logic directly into the ILP constraints on the assignment variable x_{ij}. For example, job type compatibility is enforced by two inequalities that, if $x_{ij} = 1$, force the numerical indices of the seeker's desired job type and the job's type to be equal. Similar logic applies to other compatibility checks, ensuring that x_{ij} can only be 1 if all underlying conditions are met relative to the Big-M constants.

2.5 Dissimilarity Score (d_{ij}) Calculation

The dissimilarity score d_{ij} between a seeker i and a job j is based on their respective 20-question questionnaires. It is calculated as the average absolute difference between their answers:

$$d_{ij} = \frac{\sum_{k=1}^{20} |Quest_{seeker,i,k} - Quest_{job,j,k}|}{20} \tag{14}$$

where $Quest_{seeker,i,k}$ is the k^{th} answer for seeker i and $Quest_{job,j,k}$ is the k^{th} ideal answer for job j. These scores were pre-calculated for all seeker-job pairs before solving the Part 2 model. If questionnaire lengths were mismatched (not 20 for both), a maximum dissimilarity of 5.0 was assigned as a penalty.

3 Results and Analysis

3.1 Part 1 Results: Maximizing Priority Weight

The ILP model for Part 1 was solved to maximize the total priority weight of assignments.

- **Maximum Total Priority Weight (M_W): 240.00**
- **Number of Assignments Made: 46**

This M_W value serves as the benchmark for Part 2 analysis. Detailed assignments are provided in Appendix A.

3.2 Part 2 Results: ω Experimentation

In Part 2, the model aimed to minimize the maximum dissimilarity (M) for different target levels of total priority, specified by $\omega \in \{70, 75, 80, 85, 90, 95, 100\}$ percent of M_W. The results are summarized in Table 1 and visualized in Figure 1.

Table 1: Part 2 Results: Minimizing Maximum Dissimilarity for various ω values ($M_W = 240.00$)

ω (%)	Min Max Dissim.	Num Assignments	Total Priority Achieved
70.0000	2.0000	31.0000	168.0000
75.0000	2.0000	33.0000	180.0000
80.0000	2.1000	36.0000	195.0000
85.0000	2.1500	41.0000	221.0000
90	**2.1500**	**41**	**221.00**
95.0000	2.3000	43.0000	229.0000
100.0000	2.5000	46.0000	240.0000

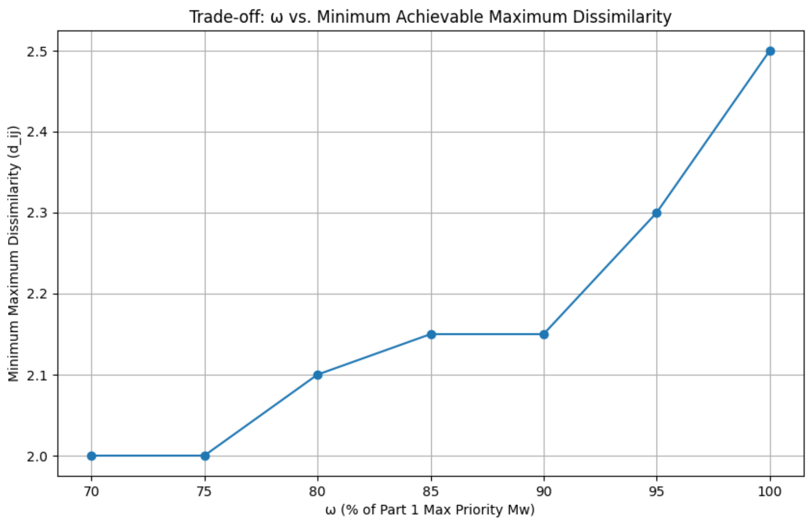

Figure 1: Trade-off: ω vs. Minimum Achievable Maximum Dissimilarity

Analysis of ω Experimentation: As observed in Table 1 and Figure 1:

- When ω is low (e.g., 70-75%), the model has more flexibility to choose matches with lower dissimilarity scores, achieving a minimum maximum dissimilarity of 2.0000. This comes at the cost of a lower total priority and fewer assignments.
- As ω increases, the constraint on achieving a higher total priority becomes tighter. This forces the model to include matches that might have higher dissimilarity scores to meet the priority target.
- The minimum maximum dissimilarity increases from 2.0000 at $\omega = 75\%$ to 2.1000 at $\omega = 80\%$, then to 2.1500 at $\omega = 85\%$. It remains at 2.1500 for $\omega = 90\%$.
- For $\omega > 90\%$, the minimum maximum dissimilarity increases more sharply (to 2.3000 for $\omega = 95\%$ and 2.5000 for $\omega = 100\%$). This indicates that to achieve these very high priority targets, some matches with relatively high dissimilarity must be accepted.
- The number of assignments generally increases with ω, as does the total priority achieved (by definition of the constraint).

The plot clearly shows a trade-off: demanding a higher total priority (higher ω) generally leads to a higher maximum dissimilarity among matches.

3.3 Choice of ω and Detailed Results

Based on the project guidelines and the observed trade-off, $\omega = \mathbf{90}\%$ was chosen for detailed analysis.

Justification for chosen $\omega = 90\%$: The project requires experimentation with ω values from 70% to 100%. A value of 90% aims to retain a significant portion of the maximum possible priority weight from Part 1 ($0.90 \times 240.00 = 216.00$), while providing flexibility to reduce the maximum dissimilarity score. Consulting Table 1 and Figure 1:

- At $\omega = 85\%$, the model achieved a maximum dissimilarity of 2.1500, with a total priority of 221.00 and 41 assignments.
- At $\omega = 90\%$, the target minimum priority is 216.00. The model still achieved a maximum dissimilarity of 2.1500, with a total priority of 221.00 and 41 assignments.
- The results for $\omega = 85\%$ and $\omega = 90\%$ are identical in terms of achieved dissimilarity, assignments, and total priority. This means the optimal solution found when targeting 85% of M_W already satisfied the 90% M_W target.
- Increasing ω further to 95% (target priority 228.00) results in a notable jump in maximum dissimilarity to 2.3000 (achieved priority 229.00).

Thus, $\omega = 90\%$ is chosen as it formally meets a higher priority target than $\omega = 85\%$ without any degradation in the maximum dissimilarity or other key metrics. It represents a point before the dissimilarity cost increases more significantly for marginal gains in total priority.

Detailed Results for $\omega = 90\%$:

- Target Minimum Total Priority: $0.90 \times 240.00 = 216.00$

- Achieved Total Priority: 221.00
- Achieved Maximum Dissimilarity: 2.1500
- Number of Assignments: 41

The model successfully found a set of 41 assignments that achieved a total priority of 221.00 (exceeding the target of 216.00), with no single assignment having a dissimilarity score greater than 2.1500. Detailed assignments for this chosen ω are provided in Appendix B.

4 Comments and Conclusion

This project successfully demonstrated the application of Integer Linear Programming to a complex job matching problem with multiple objectives and constraints.

Problem and Models:

- The problem is inherently multi-objective: maximizing job priorities (Part 1) and minimizing preference dissimilarity (Part 2) are often conflicting goals. The ω parameter in Part 2 provides a mechanism to explore this trade-off.
- The ILP models, using Big-M formulations to handle logical compatibility conditions, are powerful tools for finding optimal solutions under the given criteria. This approach directly embeds all logical conditions into the ILP structure.

Results:

- Part 1 established an upper bound for achievable total priority ($M_W = 240.00$ with 46 assignments).
- The ω experimentation in Part 2 clearly illustrated the trade-off curve. As expected, forcing the model to achieve a higher percentage of M_W restricts its choices, often leading to an increase in the maximum dissimilarity allowed for any single match.
- The choice of $\omega = 90\%$ provides a balanced solution, securing a high total priority (221.00, which is 92.08% of M_W) while keeping the maximum dissimilarity at 2.1500. This is identical to the outcome for $\omega = 85\%$, indicating that the higher priority target did not impose an additional dissimilarity cost at this level.
- Compared to $\omega = 100\%$ (max dissimilarity 2.5000, priority 240.00, 46 assignments), the $\omega = 90\%$ solution makes 5 fewer assignments (41 assignments) but significantly improves the worst-case dissimilarity (2.1500 vs 2.5000).

Conclusion: The two-stage ILP approach is effective for this job matching scenario. Part 1 provides a crucial benchmark, and Part 2 allows for a nuanced balancing of employer priorities (job weights) and candidate-job fit (dissimilarity). The choice of ω is a strategic decision. The results for $\omega = 90\%$ offer a practical and well-justified compromise, achieving a substantial portion of the maximum priority while maintaining a good level of match quality in terms of dissimilarity.

A Appendix: Part 1 Detailed Assignments

Part 1 Results: Maximize Priority-Weighted Matches
Maximum Total Priority Weight (Mw): 240.00
Number of assignments: 46

Table 2: Detailed Assignments (Seeker_ID → Job_ID) for Part 1

Seeker ID	Seeker Loc	Job ID	Job Loc	Job Type	Priority
S0003	D	J0010	B	Part-time	1.00
S0006	C	J0069	C	Contract	7.00
S0011	B	J0054	E	Part-time	2.00
S0018	E	J0052	E	Full-time	4.00
S0025	D	J0002	C	Internship	8.00
S0028	E	J0070	A	Part-time	3.00
S0029	D	J0017	B	Part-time	5.00
S0031	D	J0028	D	Part-time	6.00
S0032	F	J0068	D	Full-time	10.00
S0034	D	J0029	E	Full-time	5.00
S0038	D	J0004	B	Internship	7.00
S0061	E	J0053	E	Internship	6.00
S0065	B	J0020	E	Full-time	9.00
S0068	C	J0063	B	Contract	3.00
S0078	D	J0046	B	Full-time	4.00
S0079	B	J0003	E	Part-time	8.00
S0082	F	J0019	B	Internship	8.00
S0085	C	J0010	B	Part-time	1.00
S0088	B	J0019	B	Internship	8.00
S0089	B	J0046	B	Full-time	4.00
S0094	A	J0028	D	Part-time	6.00
S0100	E	J0019	B	Internship	8.00
S0102	F	J0068	D	Full-time	10.00
S0104	E	J0053	E	Internship	6.00
S0111	B	J0028	D	Part-time	6.00
S0121	A	J0026	D	Internship	6.00
S0128	E	J0008	F	Contract	4.00
S0132	B	J0013	B	Internship	2.00
S0151	D	J0056	C	Part-time	4.00
S0159	E	J0052	E	Full-time	4.00
S0163	E	J0069	C	Contract	7.00
S0170	C	J0052	E	Full-time	4.00
S0172	A	J0027	B	Contract	4.00
S0175	D	J0046	B	Full-time	4.00
S0176	A	J0020	E	Full-time	9.00
S0177	A	J0043	B	Part-time	3.00
S0178	F	J0053	E	Internship	6.00
S0199	F	J0056	C	Part-time	4.00

Continued on next page

Table 2: Detailed Assignments (Seeker_ID → Job_ID) for Part 1 *(continued)*

Seeker ID	Seeker Loc	Job ID	Job Loc	Job Type	Priority
S0245	E	J0070	A	Part-time	3.00
S0262	C	J0002	C	Internship	8.00
S0282	B	J0034	A	Part-time	4.00
S0291	F	J0048	D	Internship	5.00
S0330	B	J0046	B	Full-time	4.00
S0331	D	J0011	C	Full-time	2.00
S0340	C	J0046	B	Full-time	4.00
S0381	C	J0008	F	Contract	4.00

Sum of priorities in assignments (check): 240.00

B Appendix: Part 2 Detailed Assignments for $\omega = 90\%$

Part 2: Detailed Assignments for Chosen $\omega = 90\%$
Target: Achieve at least 90% of Mw (240.00), which is 216.00
Achieved Max Dissimilarity: 2.1500
Achieved Total Priority: 221.00
Number of Assignments: 41

Table 3: Detailed Assignments for Part 2, $\omega = 90\%$

Seeker ID	Seeker Loc	Job ID	Job Loc	Job Type	Priority	Dissim.
S0003	D	J0010	B	Part-time	1.00	1.6500
S0018	E	J0052	E	Full-time	4.00	2.0000
S0025	D	J0002	C	Internship	8.00	1.6500
S0028	E	J0070	A	Part-time	3.00	1.5500
S0029	D	J0017	B	Part-time	5.00	1.3000
S0031	D	J0028	D	Part-time	6.00	1.4500
S0034	D	J0029	E	Full-time	5.00	2.1500
S0038	D	J0004	B	Internship	7.00	2.0500
S0061	E	J0053	E	Internship	6.00	2.1500
S0065	B	J0020	E	Full-time	9.00	2.1500
S0068	C	J0063	B	Contract	3.00	1.4000
S0078	D	J0046	B	Full-time	4.00	1.8500
S0079	B	J0003	E	Part-time	8.00	1.7000
S0085	C	J0010	B	Part-time	1.00	1.5000
S0088	B	J0019	B	Internship	8.00	2.1000
S0089	B	J0046	B	Full-time	4.00	1.7500
S0100	E	J0019	B	Internship	8.00	2.0000
S0104	E	J0053	E	Internship	6.00	1.2000
S0121	A	J0026	D	Internship	6.00	1.4500
S0128	E	J0008	F	Contract	4.00	1.7500
S0132	B	J0013	B	Internship	2.00	2.1500
S0151	D	J0056	C	Part-time	4.00	1.7500
S0156	D	J0054	E	Part-time	2.00	1.2000
S0159	E	J0052	E	Full-time	4.00	2.0500
S0172	A	J0027	B	Contract	4.00	2.1500
S0173	E	J0069	C	Contract	7.00	1.7000
S0175	D	J0046	B	Full-time	4.00	2.1000
S0176	A	J0020	E	Full-time	9.00	1.8000
S0178	F	J0053	E	Internship	6.00	1.9500
S0199	F	J0056	C	Part-time	4.00	1.7500
S0245	E	J0070	A	Part-time	3.00	1.5500
S0253	B	J0028	D	Part-time	6.00	1.4000
S0262	C	J0002	C	Internship	8.00	1.5000
S0268	D	J0068	D	Full-time	10.00	2.1500
S0273	C	J0019	B	Internship	8.00	1.7500
S0289	E	J0069	C	Contract	7.00	1.7000

Continued on next page

Table 3: Detailed Assignments for Part 2, $\omega = 90\%$ *(continued)*

Seeker ID	Seeker Loc	Job ID	Job Loc	Job Type	Priority	Dissim.
S0308	C	J0068	D	Full-time	10.00	1.8500
S0330	B	J0046	B	Full-time	4.00	1.7000
S0340	C	J0046	B	Full-time	4.00	1.6500
S0382	C	J0028	D	Part-time	6.00	1.4000
S0398	A	J0043	B	Part-time	3.00	1.6000