FINAL PROJECT
COMPUTATIONAL METHODS FOR DATA SCIENCE
FALL SEMESTER 2023

The main goal of this final project is to apply optimization algorithms learnt from the lecture to real-life problems. In this final project, we will consider a task scheduling problem in supercomputer facility. Traditionally, there are many machines in a supercomputer facility and many users will input their multi-step jobs to the facility. Given the following conditions:

1. There are n machines, denoted as $M_1, M_2, \ldots, M_n$ in a supercomputer facility.
2. There are i jobs, denoted as $J_1, J_2, \ldots, J_i$, inputed by i different users.
3. For each job, there are j operational steps, i.e. for Job i, there are $O_{i1}, O_{i2}, \ldots, O_{ij}$ steps.
4. Each operational step O_{ij} will spend $P_{ijn} + S_{ijn}$ time to complete in Machine n, where P_{ijn} is the processing time and S_{ijn} is the setup time.
5. There is a job transfer time $T_{n_1 n_2}$ for an operational job to transferred from Machine M_{n_1} to Machine M_{n_2} when it moves from an operational step to another.

We record the processing time and setup time of each machine for every specific operational step of all jobs in a table called Task Table, and each entry of this table is in the form of P_{ijn}/S_{ijn}, i.e. the first number is the processing time and the second number is the setup time.

Below are some assumptions in this optimization problem:

1. Only one operational steps from a job can be performed in a machine in one moment (i,e, no shared machine setup).
2. The operational step cannot be stopped before competion.
3. The setup time and the delivery time can be overlapped only (i.e. processing time cannot be overlapped).
4. There is zero time to switch between jobs.
5. There is zero setup time if different steps of the same job is chosen in the same machine.

The objective is to come up with an optimal task schedule that can minimize the time for the last machine to stop (the maximum processing time among all machines). In this final project, here are a list of components that will be required.

1. **A Toy Example.** *(20 points)* Assume there is only two identical machines M_1 and M_2 and there are four jobs $J_1, \ldots, J_4$. Each job has only one operation and all setup time is assumed to be 1 for simplcity. The task table and the job transfer matrix are as follows: Find the

J_i	O_{ij}	M_1	M_2
J_1	O_{11}	3/1	3/1
J_2	O_{21}	5/1	5/1
J_3	O_{31}	9/1	9/1
J_4	O_{41}	8/1	8/1

optimal strategy by complete enumeration.

From \ To	M_1	M_2
M_1	0	1
M_2	2	0

2. **Metaheuristic Optimization Methods** *(40 points)* This part is about programming of the metaheuristic optimization for this specific question.

   (a) It is obvious that the input is the Task table and the Delay matrix, but what is the data structure or form of the output?

   (b) What is the objective function of this optimization problem?

   (c) The genetic algorithm seems to work well in this problem. Write a GA program that inputs the task table and the delay matrix, and outputs the optimal solution of your data structure.

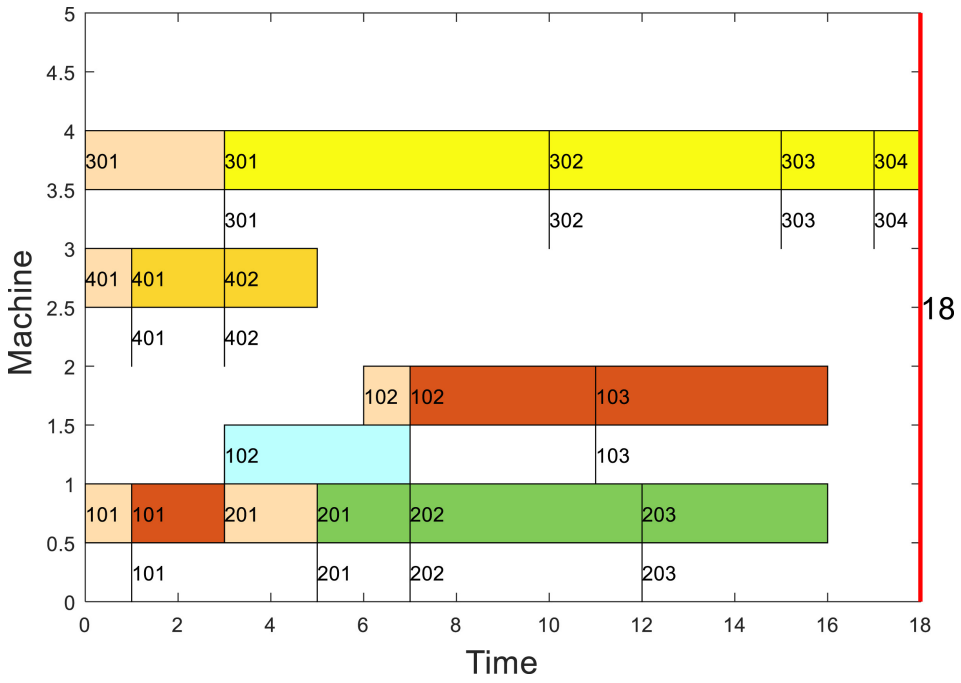
   (d) Among the remaining methods (Simulated Annealing, Tabu Search, Particle Swarm Optimization and Ant Colony Opitmization), choose two programs that you believe they work well for this problem. Do the programming with the similar inputs and otuputs.

3. **A Main Task** *(40 points)* Given the following Task Table and the Delay Matrix, find the optimal task schedule for all four machines, so that the schedule can minimize the time for the last machine to stop.

J_i	O_{ij}	M_1	M_2	M_3	M_4	M_5
J_1	O_{11}	2/1	5/3	4/1	1/1	2/2
	O_{12}	5/3	4/1	5/2	7/6	5/1
	O_{13}	4/1	5/4	5/4	4/2	5/2
J_2	O_{21}	2/2	5/4	4/1	7/5	8/7
	O_{22}	5/1	6/6	9/7	8/4	5/3
	O_{23}	4/2	5/3	4/3	54/43	5/4
J_3	O_{31}	9/4	8/7	6/4	7/3	9/9
	O_{32}	6/6	1/1	2/2	5/3	4/1
	O_{33}	2/1	5/5	4/1	2/1	4/2
	O_{34}	4/4	5/1	2/2	1/1	5/3
J_4	O_{41}	1/1	5/1	2/1	4/8	12/8
	O_{42}	5/4	1/1	2/1	1/1	2/1

From \ To	M_1	M_2	M_3	M_4	M_5
M_1	0	4	2	1	3
M_2	1	0	5	1	5
M_3	1	3	0	2	3
M_4	3	1	3	0	1
M_5	4	2	5	3	0

All optimal solutions are presented in a figure form called Gantt chart. An example of this figure (one not-optimal solution of this problem) is demonstrated as follows.



Use Job 1 as an example to illustrate. According to this task schedule, O_{11} is setup in M_1 for 1 sec, with 0 transfer time from M_1 to M_1, O_{11} is processed in M_1 for 2 sec. Then Job 1 is transferred from M_1 to M_2 that takes 4 sec, and at the same time, O_{12} is setup in M_2 for 1 sec. Then O_{12} and O_{13} are completed in M_2 for 4 sec and 5 sec respectively. There are 4 machines shown in this figure, and M_4 takes the longest processing time among all at 18 sec. This is the one value we want to minimize.

Use all three methods (GA must be included) of your choices in Q2 to get three optimal solutions, and compare their results. Please shows all details of your works.

4. **A Challenging Question.** *(Bonus 10 points)* In reality, not all operational steps can be performed in any machines of your choices, i.e. there are NA entries in Task table. Please provide a solution when such scenario happens. To demonstrate, use the same Task table where O_{23} at M_4 is changed to NA instead of 54/43, and O_{33} at M_1 is changed to NA instead of 2/1. (NOTE: You do not need to work out the whole program for demonstration, just use some figures or sentences to explain).