Team Selection Problem - Week 1

Daniel Suh

June 2018

Q: What is the current optimization problem?
We have the following parameters:

- p, a number of projects for students to choose from - in the paper, this was between 13 to 15
  side note: we probably want to know how many students can be assigned to each project, which is the motivation for the parameter c_i below

- s, a number of students to be matched - in the paper, this was approximately 75

- c, a p-length vector containing the capacity c_i of each project i - in the paper, this was not specified, and it cannot be assumed that each project takes the same number of students, i.e. 5 maximum to a project, but inspection of code may prove otherwise
  side note: making the simplifying assumption that project team size is homogeneous ($\lfloor \frac{s}{p} \rfloor$ minimum to $\lceil \frac{s}{p} \rceil$ maximum) from project to project may help for calculations, allocations, but would harm exportability

- s_j, a p-length vector containing each student j's preferences for each project i, with each entry s_{ij} ranging from 1 (least interest) to 5 (most interest).
  side note: one thing to consider about the individual s_i is that students may attempt to choose their allocations strategically in hopes of getting their first choice preference. possible behaviors may include bipolarity (5s only on the few that they want and 1s on everything else), collusion (students collaborate on what they will vote on, i.e. a select group of students mark 5s on the same few projects and below 3 for all other projects, while blacklisting the rest of the class),

- f_j, a s-length vector that accounts for each individual's preferential partners, with each entry coded as 0 (impartial to working with him/her) or 1 (desirable to working with him/her, due to complementary skills or good camaraderie) - in the paper, such a vector was not considered. f chosen for "friend".

- e_j, a s-length vector that accounts for each individual's anti-preferential partners, with each entry coded as 0 (impartial to working with him/her) or -1 (undesirable to working with him/her, due to possible previous conflict or other negative behavior) - in the paper, such a vector was considered with the limitation that at most two -1 scores. e chosen for "enemy".

- $r_j = f_j + e_j$, a s-length vector to summarize the individual's overall relationship/history-based preferences; each entry therefore takes values -1, 0, or +1. r chosen for "relationship".
- k_j, a p-length vector for each student j containing each student's background compatibility with the project, with each entry ranging from 1 (no skills match/poor grades) to 5 (skills completely match/great grades), with such background compatibility ascertained by assessing grade transcripts - in the paper, it seemed like it was possible for faculty to add their recommendations and influence the assessment of a student's capability, but such a methodology was unclear. k chosen for "knowledge".
  side note: maybe alter each student's entry by -1 to +1 point per faculty recommendation
- m_j, a p-length vector for each student j that takes into account the vectors k_j, s_j, which indicate ability to work on the project and desire to work on the project, respectively. probably a weighted average of the two. hence we may preserve the values to range from 1 to 5. m chosen for "matching".
- a_i, an approximately c_i-length vector containing the student allocations to project i, coded with the student's index $(1, 2, 3, 4, \ldots)$ or, less preferably, hashed ID number

In the paper, they used a point system as follows, but this may be modified according to our purposes/priorities:

- Understaffing/overstaffing project costs = 10000 points
- Assignment of students to projects:
  - $m_{ij} = 5$ costs 0 points
  - $m_{ij} = 4$ costs 1 point
  - $m_{ij} = 3$ costs 5 points
  - $m_{ij} = 2$ costs 1000 points
  - $m_{ij} = 1$ costs 10000 points
- Assigning against anti-preference $r_j = -1$ costs 100 points
- Assigning with preference $r_j = 1$ costs 0 points (in the paper, this was not collected)
- Assigning to neutral preference $r_j = 0$ costs 0 points (assuming that the student was neutral to working with everyone else in the class)

Q: What is the objective function?

Objective function appears to be total point cost, which we can break down into Goldilocks staffing (neither too many nor too few members to a project), correct assignment, and preferences:

$$\begin{aligned}
\text{total point cost} \quad &= \quad \text{Goldilocks staffing deviation cost} \\
&\quad\ +\text{correct assignment deviation cost} \\
&\quad\ +\text{preference deviation cost}
\end{aligned}$$

We can define the "Goldilocks staffing deviation cost" by totaling up how many students were allocated to each project i to create a vector q, and then checking whether the entry q_i is equal to or below or above the corresponding c_i

$$\text{Goldilocks staffing deviation cost} = \sum_{i \in P} [\mathbb{1}_{\{c_i = q_i\}} * 0 + \mathbb{1}_{\{c_i \neq q_i\}} * 10000] \tag{1}$$

We can define the "correct assignment deviation cost" in the same manner, by using indicator variables to check whether each student j got what they wanted or not.

$$\text{correct assignment deviation cost} = \sum_{i \in P} \sum_{j \in S} [\mathbb{1}_{\{m_{ij}=5\}} * 0 + \mathbb{1}_{\{m_{ij}=4\}} * 1 + \mathbb{1}_{\{m_{ij}=3\}} * 5 + \mathbb{1}_{\{m_{ij}=2\}} * 1000 + \mathbb{1}_{\{m_{ij}=1\}} * 10000] \tag{2}$$

We can define the "preference deviation cost" by checking for each project's student allocations whether any of the students in the group like each other or don't like each other or otherwise don't care.

$$\text{preference deviation cost} = \sum_{i \in P} \sum_{j,k \in a_i, j \neq k} [\mathbb{1}_{\{r_{jk}=-1\}} * 100 + \mathbb{1}_{\{r_{jk}=0\}} * 0 + \mathbb{1}_{\{r_{jk}=+1\}} * 0] \tag{3}$$

Simplified, our objective function becomes:

$$\begin{aligned} \text{total point cost} \quad = \quad & \sum_{i \in P} [\mathbb{1}_{\{c_i = q_i\}} * 0 + \mathbb{1}_{\{c_i \neq q_i\}} * 10000] \\ & + \sum_{i \in P} \sum_{j \in S} [\mathbb{1}_{\{m_{ij}=5\}} * 0 + \mathbb{1}_{\{m_{ij}=4\}} * 1 + \mathbb{1}_{\{m_{ij}=3\}} * 5 + \mathbb{1}_{\{m_{ij}=2\}} * 1000 + \mathbb{1}_{\{m_{ij}=1\}} * 10000] \\ & + \sum_{i \in P} \sum_{j,k \in a_i, j \neq k} [\mathbb{1}_{\{r_{jk}=-1\}} * 100 + \mathbb{1}_{\{r_{jk}=0\}} * 0 + \mathbb{1}_{\{r_{jk}=+1\}} * 0] \end{aligned}$$

We wish to minimize the objective function because we incur costs for "bad" choices - allocations that go against either the will of the faculty or the students involved. The ideal optimal is a cost of 0

Q: What are the constraints?
Constraints were defined along with the parameters above, but are summarized below.

- $1 \leq p \leq \infty$ - set by capstone program designer
- $1 \leq s \geq \infty$ - set by school
- $c_i \geq 1$ - set by capstone program sponsor
- $s_{ij} = \{1, 2, 3, 4, 5\}$ - set by students
- $f_{jk} = \{-1, 0, +1\}$ - set by students
- $e_{jk} = \{-1, 0, +1\}$ - set by students
- $r_{jk} = \{-1, 0, +1\}$ - defined by f, e above

- $k_{ij} = \{1, 2, 3, 4, 5\}$ - set by students indirectly by past performance
- $m_{ij} = x$, where $1 \leq x \leq 5$, $x \in \mathbb{R}$ - defined by k, s above