Partitioning of students into equitable groups using SolverStudio

M. Fairley*, O. Dowson
Department of Engineering Science
University of Auckland
New Zealand
*mfai035@aucklanduni.ac.nz

Abstract

Students in their final year of a Bachelor of Engineering at The University of Auckland are required to complete a course known as ENGGEN 403. As part of this course, students are partitioned into groups of around 25 students and given a single week to produce a large piece of work on a given topic. In order to make this fair and equitable, the students should be partitioned in a way to make the groups as similar as possible. This paper details the development, implementation, and results of an Excel based optimisation solution for this problem. A model was formulated to minimise a weighted combination of the greatest difference between the mean grade point average of each group, and the greatest difference between the variance of the grade point averages in each group. The model was also constrained to balance gender, ethnicity and engineering discipline. This model was then implemented in the PuLP modelling language using the SolverStudio modelling environment. The model was validated against 2013 data, and shown to improve all key metrics. The Excel spreadsheet was given to the course organiser who used the model to partition 571 students into 23 groups in the 2014 ENGGEN 403 class.

Key words: knapsack, pulp, solverstudio, group allocator

1 Introduction

Students in their final year of a Bachelors in Engineering at the University of Auckland are required to complete a course known as ENGGEN403. As part of the course, students are partitioned in to groups of around 25 students and given a single week to produce a large piece of work on a given topic. In recent years, these topics have included a rebuild plan for Christchurch, a proposal for the future of Aucklands transport network and a tender to run a new public-private partnership investment agency.

1.1 Variables of Interest

It is of interest to all involved stakeholders to make the groups as fair and equitable as possible. The variable of most interest is grade point average (GPA). There is

anecdotal evidence from students and the course organiser of groups in previous years with large numbers of high GPA students. Although this does not necessarily lead to a better grade, there is a perception of inequity amongst students from other groups. A second cause for concern was groups with a bi-modal split of high-GPA students and low-GPA students. Therefore, two key metrics in determining the quality of any partition are the difference in mean GPA between groups, and the difference in the group GPA variance between groups.

A second set of variables concern the demographic make-up of the groups. Although not as strictly related to academic performance as GPA, group demographics can have a large impact on the perceived fairness and enjoyment by students.

One key demographic variable is gender. It is widely acknowledged that women are under-represented in Engineering at the University of Auckland.

ENGGEN403 is a common paper to all the engineering disciplines at the University of Auckland. In addition, each discipline has a unique set of skills and perspective to contribute to group projects. To give all groups an equal balance of these skills, and to prevent a single discipline from dominating a group, students from each discipline should be distributed equally amongst the groups.

1.2   Previous Solution

In previous years, the job of partitioning the students into groups has been the course organisers job. The method for doing so was not rigorous, but can be considered to be a greedy heuristic.

The basic method was as follows:

Starting with the smallest discipline, distribute the students across the groups aiming for equal mean GPA. Move onto the next smallest discipline and repeat. As more students are added, try to balance gender as well as GPA. Once all students are added, perform a series of swaps between pairs of students to improve the balance.

This process usually took on the order of two days to complete.

2   Mathematical model

Sets

Students: $S = \{1, 2, \ldots, m\}$;
Groups: $G = \{1, 2, \ldots, n\}$;
Ethnicities: $E = \{$European, Māori, Pacific, Asian, Other$\}$;
Genders: $K = \{$Male, Female, Other$\}$;
Disciplines: D = $\{$Civil, Chemical & Materials, Mechatronics,Mechanical, Electrical, Software, Computer Systems, Engineering Science, Biomedical$\}$;
Students who identify as gender k: $S_k^K \subset S$;
Students who identify as ethnicity e: $S_e^E \subset S$;
Students who study discipline d: $S_d^D \subset S$.

Parameters

GPA_s = grade point average of student s;
P_s^K = gender of student s;
P_s^E = ethnicity of student s;
P_s^D = discipline of student s;
M_g = the number of students in group g.

Decision variables

x_{sg} = allocate student s to group g;
y_{gi}^I = artificial penalty variable for group g, factor i, $\quad I = \{K, D, E\}$;

G^{max} = maximum mean GPA of group;
G^{min} = minimum mean GPA of group;
V^{max} = maximum variance of group GPA;
V^{min} = minimum variance of group GPA.

Group-Allocator

The objective function is to minimise a convex combination of the spread between the maximum and minimum mean group GPA, and the spread between the maximum and minimum variance of the GPAs within a group. Adding a penalty term with a large coefficient β is necessary as some constraints are relaxed.

$$\text{Minimise } \alpha \underbrace{(G^{max} - G^{min})}_{mean} + (1-\alpha) \underbrace{(V^{max} - V^{min})}_{Variance}$$
$$+ \beta \underbrace{\sum_{g \in G} \left(\sum_{k \in K} y_{gk}^K + \sum_{d \in D} y_{gd}^D + \sum_{e \in E} y_{ge}^E \right)}_{\text{Penalty variables}} \tag{1}$$

Partitioning Constraint This constraint ensures that all students are allocated to exactly one group.

$$\sum_{g=1}^{n} x_{sg} = 1 \qquad \forall s \in S \tag{2}$$

Bounds on objective variables These constraints set the values of the variables used in the objective function.

$$\sum_{s \in S} x_{sg} \times \frac{G_s}{M_g} \geq G^{min} \qquad \forall g \in G \tag{3}$$

$$\sum_{s \in S} x_{sg} \times \frac{G_s}{M_g} \leq G^{max} \qquad \forall g \in G \tag{4}$$

$$\sum_{s \in S} x_{sg} \times \frac{G_s^2}{M_g} - \mu^2 \geq V^{min} \qquad \forall g \in G \tag{5}$$

$$\sum_{s \in S} x_{sg} \times \frac{G_s^2}{M_g} - \mu^2 \leq V^{max} \qquad \forall g \in G \tag{6}$$

Bounds on demographic variables These constraints set the minimum number of students in each group from each demographic sub-group. These constraints are relaxed via the addition of y variables.

$$\sum_{s \in S_k^K} x_{sg} + y_{gk}^K \geq \left\lfloor \frac{M_g \times |S_k^K|}{|S|} \right\rfloor \qquad \forall g \in G \quad \forall k \in K \tag{7}$$

$$\sum_{s \in S_d^D} x_{sg} + y_{gd}^D \geq \left\lfloor \frac{M_g \times |S_d^D|}{|S|} \right\rfloor \qquad \forall g \in G \quad \forall d \in D \qquad (8)$$

$$\sum_{s \in S_e^E} x_{sg} + y_{ge}^E \geq \left\lfloor \frac{M_g \times |S_e^E|}{|S|} \right\rfloor \qquad \forall g \in G \quad \forall e \in E \qquad (9)$$

Variable Bounds The allocation variables (x) are binary, all other variables are non-negative.

$$x_{sg} \in \{0, 1\}$$

$$G^{min}, G^{max}, V^{min}, V^{max}, y_{gi}^I \geq 0.$$

3   Reporting

In this section we discuss the visual aids produced to help users understand the solution.

4   Results

In this section we discuss the results of the 2013 ENGGEN403 class.

5   Discussion

Acknowledgments

References