Balanced Committee Election

June 7, 2018

In a balanced committee election, we have m candidates $C = \{1, \ldots, m\}$, n voters and a desired size $k \in \{1, \ldots, m\}$ of a committee to be elected. Each voter can select at most k candidates. This gives each candidate i a total number of votes w_i. A committee $S \subseteq C$ of size k is considered to be good if it receives many votes, i.e., if the sum of all of the votes for the candidates in S, $\sum_{i \in S} w_i$, is large. At the same time, we want to guarantee that the selected committee S is "balanced" according to important features such as gender and locality. For example, we might want S to satisfy the following two criteria:

(1) 50% male and 50% female;

(2) 80% candidates from cities and 20% candidates from the countryside.

Overall, our goal is to elect a committee $S \subseteq C$ of size k that satisfies the above criteria. Furthermore, amongst all possible committees that do so, we would like to elect the one with the most votes, i.e., with maximal $\sum_{i \in S} w_i$.

We can encode this problem mathematically by using a 0/1 variable x_i to represent whether the candidate i is selected ($x_i = 1$) or not ($x_i = 0$). To give an example, suppose that $k = 10$. Then, the requirement that the committee S is of size 10 is equivalent to saying that we must satisfy

$$\sum_{i \in C} x_i = k,$$

i.e., exactly 10 candidates must have $x_i = 1$. Suppose $P_m \subseteq C$ represents the collection of male candidates and $P_f \subseteq C$ the collection of female candidates. The gender criteria (1) is equivalent to

$$\sum_{i \in P_m} x_i = 5, \quad \text{and} \quad \sum_{i \in P_f} x_i = 5,$$

i.e., exactly 5 male candidates must have $x_i = 1$. Similarly, suppose $P_c \subseteq C$ represents the collection of candidates living in the city and $P_t \subseteq C$ the collection of candidates living in the countryside. The locality criteria (2) is equivalent to

$$\sum_{i \in P_c} x_i = 8, \quad \text{and} \quad \sum_{i \in P_t} x_i = 2.$$

If candidate i is selected, then it contributes $w_i = w_i \cdot x_i$ votes to the committee (as $x_i = 1$). Otherwise, it contributes $0 = w_i \cdot x_i$ votes (as $x_i = 0$). Thus, the total votes received by the selected committee captured by the equation

$$\sum_{i \in C} w_i \cdot x_i.$$

Using the above formulation, we can write down an "integer linear program" to encode the mathematical problem of finding the best committee that satisfies this criteria as follows:

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i \in C} w_i \cdot x_i && \text{(maximize the total votes)} \\
\text{such that} \quad & x_i \in \{0,1\} \quad \text{for all } i \in C, && \text{(0/1 variables)} \\
& \sum_{i \in C} x_i = 10, && \text{(committee size constraint)} \\
& \sum_{i \in P_m} x_i = 5, \sum_{i \in P_w} x_i = 5, && \text{(gender constraints)} \\
& \sum_{i \in P_c} x_i = 8, \sum_{i \in P_t} x_i = 2. && \text{(locality constraints)}
\end{aligned}
$$

In general, solving an integer linear program is NP-hard and hence can take a prohibitively long time. However, we can use a standard package **CPLEX** which uses a "branch-and-cut" approach to speed up the computation while still ensuring that we get the optimal solution.