

Modeling and Solving the Balanced Assignment Problem with CP Optimizer

Philippe Laborie
IBM, 9 rue de Verdun 94250 Gentilly, France
laborie@fr.ibm.com

1 The problem

This problem is described in dmcommunity.org/challenge/challenge-sep-2018. It deals with assignment of people to certain activities, for instance: assigning students to project groups or assigning professors to offices. Let's assume we have 210 people that belong to 4 categories as described in the file `BalancedAssignment.csv`:

NAME	DEPARTMENT	LOCATION	GENDER	TITLE
BIW	NNE	Peoria	M	Assistant
KRS	WSW	Springfield	F	Assistant
TLR	NNW	Peoria	F	Adjunct
VAA	NNW	Peoria	M	Deputy
JRT	NNE	Springfield	M	Deputy
...				

You need to assign these 210 people to 12 groups of 16, 17, 18, or 19 people in every group. Of course, a person may be assigned to only one group. The objective is to maximize the diversity of people in the groups. *Diversity* means that it is preferable to assign people with the same characteristics to different groups. In this note we use a definition of *diversity* that is using a simple rule that measures the penalty when two people with similar characteristics are assigned to the same group and then minimize the total penalty. More precisely, the penalty for having two people with similar characteristics in the same group is equal to the number of characteristics they have in common. For instance if BIW and JRT are in the same group, it will incur a penalty of 2 because they have two characteristics in common: their department (NNE) and their gender (M). Thus the problem is to assign the people to the different groups while minimizing the total penalty.

2 Formal definition

Let N denote the set of people and C the set of categories (here: DEPARTEMENT, LOCATION, GENDER, TITLE). For a person $i \in N$ and a characteristic $c \in C$, the value of this characteristic is denoted t_{ik} . The penalty for having two persons i and j in the same group is: $P_{ij} = |\{k \in C : t_{ik} = t_{jk}\}|$. Let G denote the set of groups, A (resp. B) the minimal (resp. maximal) number of people in a group. The problem can be defined as follows:

$$\begin{aligned}
& \min \sum_{i,j \in N, i < j} P_{ij} \cdot (x_i == x_j) \\
& A \leq \text{count}([x_i]_{i \in N}, g) \leq B \quad \forall g \in G \\
& \text{integer } x_i \in G \quad \forall i \in N
\end{aligned}$$

Where $\text{count}([x_i]_{i \in N}, g)$ returns the number of variables x_i for $i \in N$ such that $x_i = g$ that is, the number of people in group g . Note that for a pair of persons $\{i, j\}$ in the same group, we apply the penalty only once (and not for both (i, j) and (j, i) which would result in a total penalty twice as big).

3 Complexity

The Balanced Assignment Problem is **NP-Complete** in the strong sense. The decision version of the problem is clearly in NP. We show NP-Completeness by a transformation from graph K-colorability.

Let $G = (V, E)$ and $K \leq |V|$ be an instance of graph K-colorability problem with $V = \{1, \dots, n\}$. We can transform this instance into an instance of Balanced Assignment Problem with K groups (one for each color) where the problem is to decide whether there exists a solution of total penalty 0. Each vertex $i \in V$ will be a person. For each edge $(i, j) \in E$, we define a category C_{ij} as follow:

$$\begin{aligned}
C_{ij}(i) &= -1 \\
C_{ij}(j) &= -1 \\
C_{ij}(k) &= k \quad \forall k \in V, k \neq i, k \neq j
\end{aligned}$$

As the total penalty needs to be 0, the penalty associated with each of the K groups needs to be 0. It means that two persons i, j such that $(i, j) \in E$ cannot belong to the same group (same color) as otherwise, the penalty due to having them sharing the same characteristic in the same group ($C_{ij} = -1$) would be 1. So any solution to the Balanced Assignment Problem is a valid K-coloring of the graph.

Reciprocally, any solution to the K-coloring problem will be such that for each pair (k, l) of elements in a group (same color), and for every category C_{ij} , we necessarily have $\{i, j\} \neq \{k, l\}$ and thus $C_{ij}(k) \neq C_{ij}(l)$. This means that the total penalty will be 0 and the solution is also a solution to the associated Balanced Assignment Problem.

Though the problem is NP-Complete, the above transformation from graph K-colorability requires a lot of categories in the corresponding Balanced Assignment Problem (at most n^2 as we create one category per edge in the graph). In practice, I think that when the problem uses few categories (4 in the example), then the problem is not very difficult and very good solution can be build using simple heuristics suggested by the computation of lower bounds in section 6.

4 A CP Optimizer formulation

The complete CP Optimizer formulation of the problem in Python is given on Fig. 1. For more information on CP Optimizer, we refer interested readers to [1].

5 A solution

Using this model we can find a solution with total penalty **3370** shown on Fig. 2. We will see in the next section that this solution is **optimal** because we show, using some simple mathematical considerations, that a lower bound on the total penalty is 3370.

Figure 1: CP Optimizer model in Python

```

1 # 1. READING THE DATA
2
3 import csv
4 with open('BalancedAssignment.csv') as csvfile:
5     t = [p[1:] for p in csv.reader(csvfile)]
6
7 G = range(12)           # Groups
8 N = range(len(t))       # People
9 C = range(len(t[0]))    # Categories
10 A,B = 16,19             # Bounds on groups size
11
12 # Penalty function
13 def P(i,j): return sum((t[i][c]==t[j][c]) for c in C)
14
15 # 2. MODELING THE PROBLEM WITH CP-OPTIMIZER
16
17 from docplex.cp.model import *
18 m = CpoModel()
19
20 # Decision variables: x[i] is the group of person i
21 x = [integer_var(G) for i in N]
22
23 # Objective: minimize total penalty
24 m.add(minimize(sum(P(i,j)*(x[i]==x[j]) for i in N for j in range(i))))
25
26 # Group size constraints
27 m.add([in_range(count(x,g), A, B) for g in G])
28
29 # 3. SOLVING THE PROBLEM
30
31 sol = m.solve(trace_log=True)

```

The distribution of people characteristics in this solution is shown on the plots on Fig. 3 and 4.

6 Computing a lower bound

Let S denote the set of people characteristics (for instance a characteristic can be DEPARTMENT=NNE or GENDER=M). We denote $h_{is} = 1$ if person $i \in N$ has characteristic $s \in S$, $h_{is} = 0$ otherwise. Given two persons i and j the penalty for having them in the same group is the number of characteristics they have in common, that is:

$$P_{ij} = \sum_{s \in S} h_{is} \cdot h_{js}$$

Suppose we have a feasible solution. For a group $g \in G$, we denote E_g the set of persons in group g . The total penalty of the solution is:

Figure 2: Details of an optimal solution

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
1	DAN	DHY	TLR	MJR	MES	VAA	KRS	BIW	DCN	CWT	DBN	JAD
2	SWR	CWU	LLR	HCN	JRS	LVO	AMR	JRT	ECA	AJH	REN	JCO
3	FEE	OGZ	TAR	TBG	FTR	EAZ	DWO	CAE	MRL	RFS	FWI	RVY
4	RJY	CAH	CGH	MAI	SJO	JDS	DHE	CPR	PWS	PMO	WMS	PSN
5	DHR	KWY	TEN	JWS	RGR	GCY	RPR	CAR	LVN	ADS	SWS	MKN
6	HJO	FJT	RCG	CFS	MPO	EJD	IGN	AGI	KWR	JMS	WAD	SCS
7	SBK	KKA	KKI	HFE	PMY	RMN	JAE	ACI	RLS	WLB	RCR	CTS
8	MBM	VMA	CGY	DGR	AVI	RHD	SBM	DLT	ABE	TWD	PSO	DET
9	BCG	NMA	FJL	RLN	MTH	SME	HKA	NHI	SCR	HMN	RHE	TMN
10	BLE	MMY	SDN	PSL	AJS	RDN	MRH	RDL	DSF	LSR	MRE	SGA
11	JWB	SMA	GDE	WFS	RHA	DMN	JMQ	SJH	MLE	ANV	MNA	EGR
12	BAN	LMN	MRZ	MKA	RJZ	RJE	CMY	KLN	JMA	SNN	MBY	CEH
13	MBT	ESO	MMH	API	PST	SML	SOL	APT	JMT	EGL	HAN	MMA
14	MFR	UJT	PSZ	JPY	KRA	JJR	PAS	DJA	MSR	GMR	CSN	AJT
15	MDI	JAP	ETN	DLD	TJY	PSA	GGH	RGF	RSR	ASO	EHR	RSN
16	HWG	JSG	REZ	TBR	SAS	TTI	AJU	PSY	RGE	SER	KWN	ASA
17	JLN	JWR	XYF	LHS	BLW	MET	MRK	CGS	GWZ	FSM	HRR	MSL
18	HMG	AWN		WHF			RMS		JEN	SCG		

$$\begin{aligned}
\sum_{g \in G} \sum_{i,j \in E_g, i < j} P_{ij} &= \sum_{g \in G} \sum_{i,j \in E_g, i < j} \sum_{s \in S} h_{is} \cdot h_{js} \\
&= \sum_{s \in S} \sum_{g \in G} \sum_{i,j \in E_g, i < j} h_{is} \cdot h_{js} \\
&= \sum_{s \in S} \sum_{g \in G} \frac{1}{2} \cdot \sum_{i \in E_g} h_{is} \cdot \left(\sum_{i \in E_g} h_{is} - 1 \right) \\
&= \frac{1}{2} \sum_{s \in S} \sum_{g \in G} n_{gs} (n_{gs} - 1)
\end{aligned}$$

Where $n_{gs} = \sum_{i \in E_g} h_{is}$ denotes the number of persons with characteristic s allocated to group g .

We see that the objective function is separable by characteristics $s \in S$. To compute a lower bound, we will consider $|S|$ independent problems with only the people with characteristic s . These $N_s = \sum_{i \in N} h_{is}$ persons have to be partitioned into $|G|$ groups so as to minimize $\sum_{g \in G} n_{gs} (n_{gs} - 1)$. If we relax the integrality on n_{gs} , a trivial optimal solution is to equilibrate the number of people in each group, so taking $n_{gs} = N_s / |G|$. Thus a valid lower bound for the original problem is :

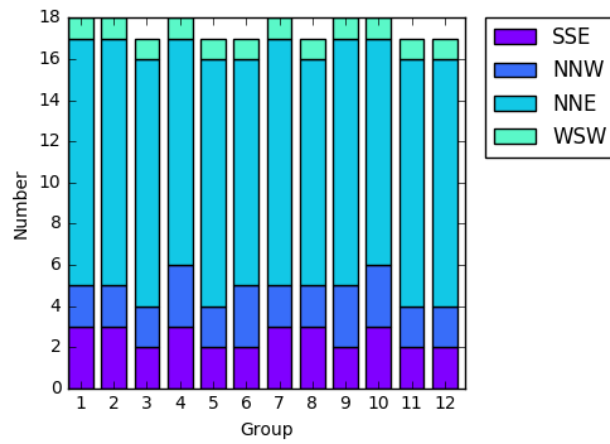
$$\frac{1}{2} \cdot \sum_{s \in S} \frac{N_s}{|G|} \cdot \left(\frac{N_s}{|G|} - 1 \right)$$

In fact it is easy to show that an optimal solution for characteristic s is to have $|G| \cdot (\lfloor \frac{N_s}{|G|} \rfloor + 1) - N_s$ groups with $\lfloor \frac{N_s}{|G|} \rfloor$ persons and $N_s - |G| \cdot \lfloor \frac{N_s}{|G|} \rfloor$ groups with $\lfloor \frac{N_s}{|G|} \rfloor + 1$ persons. This gives a total penalty for characteristic s equal to:

$$\lfloor \frac{N_s}{|G|} \rfloor \cdot (2 \cdot N_s - |G| \cdot (\frac{N_s}{|G|} - 1))$$

Figure 3: Distribution of *department* and *location* among groups

Distribution of employees in each group by department



Distribution of employees in each group by location

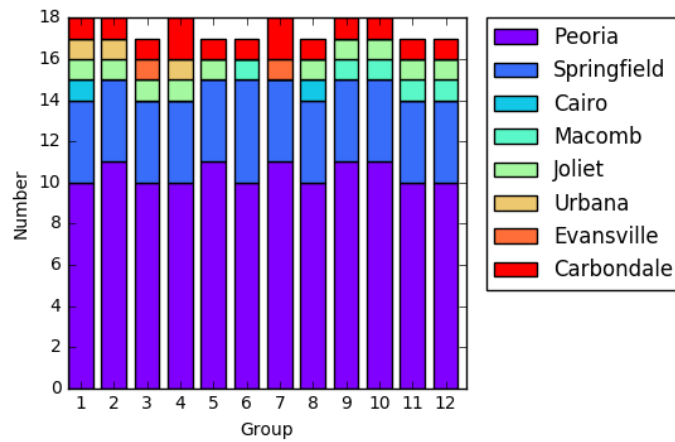
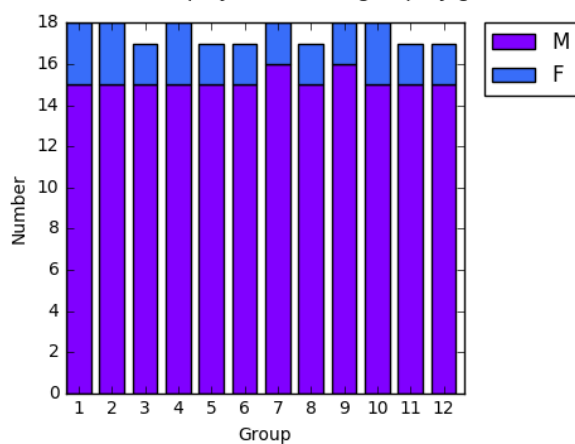
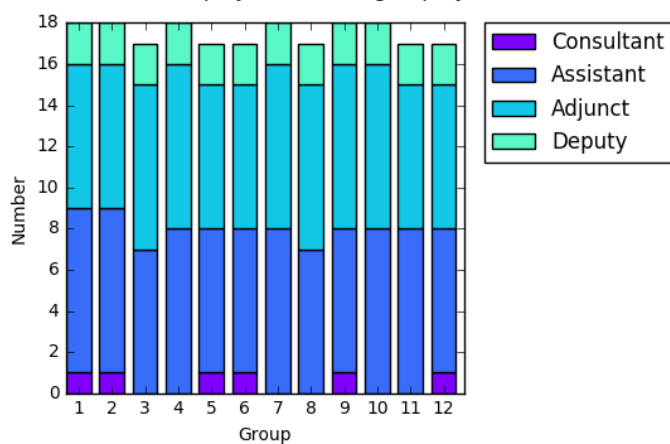


Figure 4: Distribution of *gender* and *title* among groups

Distribution of employees in each group by gender



Distribution of employees in each group by title



Thus, a valid lower bound for the original problem is :

$$\sum_{s \in S} \lfloor \frac{N_s}{|G|} \rfloor \cdot (2 \cdot N_s - |G| \cdot (\frac{N_s}{|G|} - 1))$$

If we compute it on the problem instance it gives a lower bound of **3370**.

References

- [1] Philippe Laborie, Jérôme Rogerie, Paul Shaw, and Petr Vilím. IBM ILOG CP Optimizer for Scheduling. *Constraints Journal*, 23(2):210–250, 2018.