PARAMS

- k_i instance performance indicator
- K_s average service performance indicator
- z_i k_i / k_s instance performance ratio
- S instance shutdown threshold
- I_a set of active instances (not booting)
- n $|I_a|$
- P_i previous weight of instance i
- b # of booting instances

VARIABLES

- W_i weight of instance i ($W_i \in [0,1]$)
- a_i $\begin{cases} 0 & \text{if instance } i \text{ must be shutdown} \\ 1 & \text{otherwise} \end{cases}$

OBJ FUNC

$$\min \sum_{i \in I_o} \underbrace{\frac{1}{z_i}}_{\text{Se } i \text{ è buona } z_i > 1 \text{ quindi conviene avere } W_i \text{ alto}} W_i - \underbrace{z_i\, Q_i}_{\text{Se l'istanza resta a casa conviene, scalato per } z_i}$$

Se i è buona
$z_i > 1$
quindi conviene
avere W_i alto

Se l'istanza
resta a casa
conviene,
scalato per
z_i

CONSTRAINT

1 $$W_i \leq Q_i$$

2 $$W_i \geq \frac{S}{n} Q_i$$

3 $$\sum_{i \in I_o} W_i = 1 - S/n$$

4 $$W_i \leq \frac{k_i}{k_j} W_j + (1 - Q_j)$$ $k_i \geq k_j$

5 $$W_i \leq z_i P_i + z_i \sum_{i \neq j} P_j (1 - Q_j)$$

1. To impose that an instance to shutdown has $W_i = 0$

2. To impose that an active instance should have a W_i higher than a certain threshold S.

S is the ratio between the

# of requests processed by the instance and the

# of requsts that an instance should process in an ideal case (when the load is distributed equally)

L'idea è "spegni le istanze che si allontanano molto dal caso ideale. Se è così, è perché hanno performance molto basse".

Quindi ($r_i \rightarrow$ # richieste istanza i)

$$S = \frac{r_i}{\sum r_i / n} \rightarrow \text{caso ideale}$$

Ma S non va calcolato, va visto come percentuale

Tipo, $S = 0{,}8$ indica che l'istanza i è lontana dal caso ideale del 20%

Ma $\frac{r_i}{\sum r_i} = W_i$, quindi il W_i soglia è $= S \cdot n$

Quindi vogliamo $W_i \geq S/n$

③ The weights should have unitary sum

④ We want that instance weights are proportional to their performances, i.e. that $W_i = \frac{k_i}{k_j} W_j$.

However, if $W_j = 0$, it doesn't mean that W_i must be $= 0$ if $k_i > k_j$

Hence, we want

$$W_i \leq \frac{k_i}{k_j} W_j + (1 - a_j) \text{ if } k_i \geq k_j,$$ so that if $a_j = 0$, W_i can be for sure ≤ 1

⑤ We want that a weight should not grow too much with respect to its previous weight if there are no instances to shut down. If there are, constraint ④ will take care of splitting the load fairly