Star clustering for large networks with VNS

N. Mladenović

LAMIH, University of Valenciennes

Nizni Novgorod, November 2014

Problem description - SPLP

- given a set of customers C
- given a set of possible facilities F
- for each possible facilities f it is given fix cost fc_f of opening facility f
- for each pair (customer, facility) (c, f) it is known a cost c_{cf} of servicing customer c at facility f
- it is necessary to determine set of opened facility and for each customer c a corresponding opened facility $f(c)$ such that sum of fixed cost and costs of servicing is minimized

Problem description, p-median

- given a set of locations L
- given a set of possible "centers" F
- for each pair (location, center) (l, c) it is known a distance c_{lc} between location l and center c
- it is necessary to determine set of p centers such that sum of distances between locations and nearest centers is minimized

Mathematical programming formulation, SPLP

$$\min \sum_{f \in F} fc_f y_f + \sum_{c \in C} \sum_{f \in F} c_{cf} x_{cf} \tag{1}$$

such that

$$\sum_{f \in F} x_{cf} = 1 \qquad \forall c \in C \tag{2}$$

$$x_{cf} \leq y_f \qquad \forall c \in C, \forall f \in F \tag{3}$$

$$y_f, x_{cf} \in \{0, 1\} \tag{4}$$

Mathematical programming formulation, p–median

$$\min \sum_{l \in L} \sum_{c \in C} c_{lc} x_{lc} \tag{5}$$

such that

$$\sum_{c \in C} y_c = p \tag{6}$$

$$\sum_{c \in C} x_{lc} = 1 \qquad \forall l \in L \tag{7}$$

$$x_{lc} \leq y_c \qquad \forall l \in L, \forall c \in C \qquad y_c, x_{lc} \in \{0, 1\} \tag{8}$$

VNS - Representation of solution

- Each solution can be represented as permutation x of numbers $\{1, 2, 3, ..., m = |F|\}$ where
  - first p elements are labels of opened facilities
  - remaining $m - p$ elements are indices of closed facilities
- for each customer $i \in C$ it is determined
  - c_i' - label of opened facility with smallest servicing cost (*first center*).
  - c_i'' - label of facility with second smallest servicing cost (*second center*).

Neighborhoods for Local Search

- Drop - contains solution obtained by closing one of opened facilities
- Add - contains solution obtained by opening one of currently closed facilities
- Interchange (Swap) - contains solution obtained by closing one opened and opening one of currently closed facility

Determining Best Out

- Lat us denote with x_o set of opened facilities in current solution
- for fixed facility $f' \notin x_o$ and each facility $f'' \in x_o$ it possible calculate difference between objective value for set of $x_o \setminus \{f''\} \cup \{f'\}$ and objective value for set x_o (denote this difference with $\Delta_{f'f''}$).
- facility $f'' \in x_o$ having minimal value of $\Delta_{f'f''}$ is best choice for closing and we will call this facility *Best out*

Determining Best out

- For each customer c perform the following steps
  - If $c_{cf'} \leq c_{cc'_c}$ then add difference $c_{cf'} - c_{cc'_c}$ to $\Delta^{(1)}$
  - If $c_{cf'} > c_{c'_c f}$ then add $\min\{c_{cf'} - c_{cc'_c}, c_{cc''_c} - c_{cc'_c}$ to $\Delta^{(2)}_{c'_c}$
- Determine open facility f'' having smallest value of $\Delta^{(2)}_{f''} - fc_{f''}$.
- Facility f'' is Best out and total change of objective value is

$$\Delta^{(1)} + \Delta^{(2)}_{f''} + fc_{f'} - fc_{f''}$$

- The value $\Delta^{(1)} + fc_{f'}$ is total change of objective value obtained by opening facility f' while
- the value $\Delta^{(2)}_{f''} - fc_{f''}$ total change of objective value after closing facility f''.

Shakinge

Shaking in neighborhood N_k consists of k steps of the following steps:

- Drop move with probability 0.2
- Add move with probability 0.2
- Swap move with probability 0.6

Variable Neighborhood Decomposition Search

In order to speed up method we develop VNDS by creating subproblems in the following way:

- choose at random one open facility f
- determine $k-1$ nearest opened facilities
- Select all customers whose first center is one of k selected facilities

p-median

- VNS for *p*-median is implemented in similar way.
- Local search consists only of Swap move (because number of centers is fixed to *p*)
- fix costs does not exists in this case

Clustering data by using *p*-median

We try to apply *p*-median in order to cluster clients of our mobile network.
Clients are represented by two coordinates (information) - number of sent SMS and total length of all calls during 3 months.