Stručni rad
Prihvaćeno 13. 04. 2011.

Kristian Sabo
Sanja Scitovski

Lokacija objekata u ravnini*

Location of Objects in a Plane

ABSTRACT

In the paper we consider a direct and the inverse location problem in the plane. Thereby we use different distance-like functions with appropriate illustrations. Several examples from various areas of applications are given.

Key words: data clustering, location problem, k-means, k-median, optimization

MSC 2010: 62H30, 68T10, 90C26, 90C27, 91C20, 47N10

Lokacija objekata u ravnini

SAŽETAK

U radu razmatramo izravni i obratni problem lokacije objekata u ravnini uz korištenje različitih kvazimetričkih funkcija s odgovarajućim ilustracijama. Dano je nekoliko primjera iz različitih područja primjena.

Ključne riječi: grupiranje podataka, klasteri, problem lokacije, k-sredina, k-median, optimizacija

1 Uvod

U radu razmatramo sljedeći problem lokacije: Uz pretpostavku da su poznate lokacije objekata $c_1,\dots,c_k$ u ravnini, dani skup točaka $S = \{a_i = (x_i,y_i)\colon i = 1,\dots,m\}$, $(m \gg k)$ treba grupirati (razdijeliti) na k nepraznih disjunktnih skupova (klastera) $\pi_1,\dots,\pi_k$, tako da j-tom klasteru π_j pripadnu one točke koje su u nekom smislu najbliže j-tom centru c_j. Alternativno, ovaj problem mogao bi se postaviti i na drugi način. Uz pretpostavku da je poznata particija $\Pi = \{\pi_1,\dots,\pi_k\}$ sastavljena od nepraznih disjunktnih podskupova (klastera) zadanog skupa $S \subset \mathbb{R}^2$, treba odrediti centre $c_1,\dots,c_k \in \mathbb{R}^2$ klastera $\pi_1,\dots,\pi_k$.

Odgovarajući obratni problem bio bi traženje optimalne lokacije centara $c_1^\star,\dots,c_k^\star \in \mathbb{R}^2$, na osnovi kojih bi mogli definirati i odgovarajuće optimalne klastere $\pi_1^\star,\dots,\pi_k^\star$.

Broj svih particija m-članog skupa S sastavljenih od nepraznih disjunktnih skupova $\pi_1,\dots,\pi_k$ jednak je Stirlingovom broju druge vrste $\left\{ {m \atop k} \right\}$ (vidi [10], str. 257 ili [22]), gdje je

$$\left\{ {m \atop k} \right\} = \frac{1}{k!}\sum_{j=1}^{k}(-1)^{k-j}\binom{k}{j}j^m,$$

koji u praktičnim primjenama može biti izuzetno velik (vidi [17]). Problem traženja optimalne particije spada u NP-teške probleme (vidi [7]) nekonveksne optimizacije općenito nediferencijabilne funkcije više varijabli, koja najčešće posjeduje značajan broj stacionarnih točaka.

Postoji opsežna recentna literatura iz ovog područja s različitim i brojnim primjenama, a može se pronaći pod nazivom *problem k-sredina* ili *problem k-medijana*[1] [11, 12, 16–18]. Tako se mogu pronaći brojne primjene u poljoprivredi (primjerice, razvrstavanje oranica prema plodnosti zemljišta), biologiji (primjerice, klasifikacija kukaca u grupe), genetici, medicini, prometu, kod biranja lokacije građevinskih objekata, kod razumijevanja klimatskih kretanja, u upravljanju (primjerice, rangiranje gradova i općina za potrebe financijske podrške), u poslovanju, u društvenim i humanističkim znanostima itd. Navedimo nekoliko konkretnih primjera:

Primjer 1. (Primjena u potresnom inženjerstvu)
Metode grupiranja podataka često se koriste u različitim inženjerskim primjenama. Navedimo jednu primjenu u za život važnom potresnom inženjerstvu. Na osnovi seizmoloških podataka iz prethodnog razdoblja klasterskom analizom moguće je procijeniti mjesta nepouzdana za lokaciju građevinskih objekata. U radu [21] prikazana je primjena klasterske analize u procjenjivanju oštećenja cjevovoda (voda, nafta, plin) i prepoznavanju područja nastajanja visokih oštećenja na njima. Ta mjesta uglavnom su problematična područja (seizmički kritična) i visoko rizična u nastaja nju nedostataka i kvarova na cjevovodima. Razumijevanje razloga nastajanja oštećenja na tim mjestima može poboljšati prevenciju i ublažavanje oštećenja cjevovoda.

[1] engl. k-means problem, k-median problem

* Rad je napisan uz potporu Ministarstva znanosti, tehnologije i športa Republike Hrvatske u okviru projekta 235-2352818-1034.

Primjer 2. (Pretraživanje teksta)

Postoji opsežna literatura o primjeni klasterske analize kod pretraživanja teksta (vidi primjerice [7, 14]), pri čemu uzorak od jedne ili više riječi treba pronaći u nekom dokumentu. U ovom slučaju općenito se radi o izuzetno velikim skupovima podataka visoke dimenzije, a rezultat pretrage očekuje se u realnom vremenu.

Primjer 3.

(Detekcija opasnih mjesta u prometu)

U doktorskoj dizertaciji [11] navodi se primjer primjene klasterske analize kod detekcije opasnih mjesta na autocesti "New Jersey Turnpike". Na osnovi podataka o mjestima i vrsti prometnih nezgoda detektiraju se kritične lokacije.

Primjer 4. (Definiranje izbornih jedinica)

Pretpostavimo da su poznati podaci (a_i, w_i), $i = 1, \ldots, n$ *o lokacijama naselja* a_i *s brojem glasača* w_i. *Treba odrediti izborne jedinice* $\pi_1, \ldots, \pi_k$, *tako da za svaku izbornu jedinicu* π_j *vrijedi*

(*i*) $(1-p)\frac{m}{k} \leq |\pi_j| \leq (1+p)\frac{m}{k}$,

(*ii*) $d(c_j, a_i) \leq r, \quad \forall a_i \in \pi_j$,

gdje su $c_1, \ldots, c_k$ *centri izbornih jedinica,* $m = \sum_{i=1}^{n} w_i$, *a* $p > 0$ *i* $r > 0$ *su zadani brojevi. Uvjet* (*i*) *osigurava ravnomjernost broja glasača po izbornim jedinicama do na p%, a uvjet* (*ii*) *definira maksimalnu udaljenost centra* c_j *izborne jedinice* π_j *do naselja u toj izbornoj jedinici.*

Primjer 5.

Promatramo skup korisnika koje na neki način treba povezati s još neizgrađenim objektima kao što su primjerice željezničke stanice, sportski kompleksi, knjižnice, antene ili supermarketi. Ovdje se prirodno pojavljuje problem određivanja optimalne lokacije objekata, tako da objekti u nekom smislu budu što bliže korisnicima. Osim toga dodatno možemo zahtijevati da troškovi povezivanja budu primjereno raspodijeljeni na sve korisnike te isto tako da prihodi koje ostvaruju objekti budu primjereno rasopodijeljeni na sve objekte i na taj način svi korisnici kao i objekti budu zadovoljni. U tom slučaju navedeni optimizacijski problem može se promatrati kao specijalni oblik problema lokacije, koji je u engleskom govornom području poznat pod nazivom "facility location games" *(vidi primjerice [3, 9]).*

2 Izravni problem lokacije

Funkciju $d\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_+$, koja ima barem svojstvo pozitivne definitnosti

$$d(x,y) \geq 0 \quad \& \quad d(x,y) = 0 \Leftrightarrow x = y,$$

i pomoću koje se lako računa centar c svakog diskretnog skupa $\pi \subset \mathbb{R}^2$

$$c = \underset{z \in \mathrm{conv}(\pi)}{\operatorname{argmin}} \sum_{a \in \pi} d(z,a), \tag{1}$$

zovemo *kvazimetrička funkcija*[2] (vidi primjerice [13, 20]). Najpoznatija kvazimetrička funkcija je tzv. *kvazimetrička funkcija najmanjih kvadrata* $d_{LS}(a,b) = \|a-b\|_2^2$, $a, b \in \mathbb{R}^2$, koja osim svojstva pozitivne definitnosti ima i svojstvo simetričnosti, ali ne zadovoljava nejednakosti trokuta. Centar c diskretnog skupa $\pi \subset \mathbb{R}^2$ u ovom slučaju je centroid $c = \frac{1}{|\pi|} \sum\limits_{a \in \pi} a$ (u fizici i mehanici težište ili Steinerova točka) skupa π.

Za dani skup točaka $S = \{a_i = (x_i, y_i) \in \mathbb{R}^2 \colon i = 1, \ldots, m\} \subset \mathbb{R}^2$ izravni problem lokacije možemo definirati na dva načina:

A: Ako su poznati centri $c_1, \ldots, c_k \in \mathbb{R}^2$ $(k \ll m)$, treba odrediti particiju skupa S sastavljenu od nepraznih disjunktnih skupova $\pi_1, \ldots, \pi_k$, tako da bude

$$a_i \in \pi_j \Leftrightarrow d(c_j, a_i) \leq d(c_s, a_i), \quad \forall s = 1, \ldots, k; \tag{2}$$

B: Ako je poznata particija $\Pi = \{\pi_1, \ldots, \pi_k\}$ skupa S sastavljena od nepraznih disjunktnih skupova, treba odrediti centre $c_1, \ldots, c_k \in \mathbb{R}^2$, $(k \ll m)$ klastera prema (1).

Primjerice ako su u Primjeru 4 poznati centri $c_1, \ldots, c_k$ izbornih jedinica, a treba definirati izborne jedinice $\pi_1, \ldots, \pi_k$, radi se o rješavanju izravnog problema A. Korištenjem različitih kvazimetričkh funkcija dobivaju se razni principi – kriteriji grupiranja. Osim spomenute kvazimetričke funkcije najmanjih kvadrata d_{LS}, u literaturi [2, 4, 7, 11, 13, 15, 17] često se koristi

[2]Oznaka $v = \underset{u \in D}{\operatorname{argmin}} f(u)$ znači da funkcija f u točki $v \in D$ postiže svoju najmanju vrijednost na skupu D

(*i*) $d_1(a,b) = \|a-b\|_1 = |x_1-x_2| + |y_1-y_2|,$
$a = (x_1,y_1),\, b = (x_2,y_2)$
(Manhattan ili taxicab udaljenost)

(*ii*) $d_\infty(a,b) = \|a-b\|_\infty = \max\{|x_1-x_2|, |y_1-y_2|\},$
$a = (x_1,y_1),\, b = (x_2,y_2)$
(Čebiševljeva udaljenost)

(*iii*) $d_B(a,b) = x_1 \ln \frac{x_1}{x_2} + y_1 \ln \frac{y_1}{y_2} - x_1 - y_1 + x_2 + y_2,$
$a = (x_1,y_1),\, b = (x_2,y_2)$
(Bregmanova generalizirana I-udaljenost ili Kullbach-Leiblerova udaljenost).

Spomenimo da kvazimetričke funkcije d_1 i d_∞ zadovoljavaju još dodatno svojstva simetričnosti i nejednakosti trokuta te su zbog toga prave metričke funkcije, za razliku od Bregmanove generalizirane I-udaljenosti d_B za koju se lako vidi da ne zadovoljava niti jedno od spomenutih svojstava metrike. Navedena Bregmanova generalizirana I-udaljenosti d_B samo je jedna iz klase kvazimetričkih funkcija koje su u literaturi poznate kao Bregmanove udaljenosti. Važno svojstvo Bregmanovih udaljenosti je lako računanje centra diskretnog skupa u smislu formule (1), a za mnoge ovakve kvazimetričke funkcije za centar diskretnog skupa mogu se dobiti i eksplicitne formule (vidi [13]). U literaturi postoji veliki broj različitih primjena klasterske analize koje se zasnivaju na korištenju Bregmanovih udaljenosti kao što su primjerice teorija informacija, klasifikacija teksta, obrada signala, analiza govora itd. Više o tome se može naći u [1].

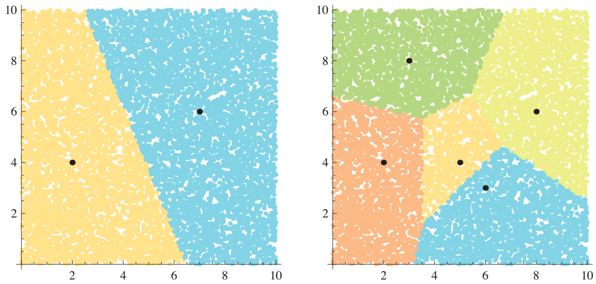
2.1 Izravni problem lokacije uz d_{LS} kvazimetričku funkciju

Za dani skup točaka $S \subset \mathbb{R}^2$ potražit ćemo rješenje izravnog problema lokacije *A* i izravnog problema *B* uz korištenje d_{LS} kvazimetričke funkcije. Ako su poznati centri $c_1,\ldots,c_k \in \mathbb{R}^2$, particiju $\Pi = \{\pi_1,\ldots,\pi_k\}$ određujemo *principom minimalnih udaljenosti* tako da bude (vidi [7, 13, 17, 19])

$$a_i \in \pi_j \Leftrightarrow \|c_j - a_i\|_2 \le \|c_s - a_i\|_2, \forall s = 1,\ldots,k. \tag{3}$$

Za $k = 2$ ovaj problem (vidi [19]) svodi se na određivanje simetrale dužine $\overline{c_1c_2}$, a može se eksplicitno riješiti kao što je prikazano na Slici 1.

a) $m = 10\,000$, $k = 2$ b) $m = 10\,000$, $k = 5$



Slika 1: *Grupiranje skupa S u klastere*

Pokazuje se da u općem slučaju ovaj problem vodi na konstrukciju Voronoijevih dijagrama (vidi [7, 13, 15, 17]). O konstrukciji Voronoijevih dijagrama vidi također [8]. Na Slici 1 ovaj problem ilustriran je primjenom vlastitog `Mathematica-modula` za $m = 10\,000$ podataka u ravnini i $k = 2$, odnosno $k = 5$ centara.

Ako je poznata particija $\Pi = \{\pi_1,\ldots,\pi_k\}$ skupa $S = \{a_i = (x_i,y_i) \in \mathbb{R}^2 : i = 1,\ldots,m\} \subset \mathbb{R}^2$, centre klastera $c_1,\ldots,c_k$ dobivamo na sljedeći način (vidi [17])

$$c_j = \underset{z \in \mathrm{conv}(\pi_j)}{\mathrm{argmin}} \sum_{a \in \pi_j} \|z - a\|_2^2 = \left(\frac{1}{|\pi_j|} \sum_{a_i \in \pi_j} x_i, \frac{1}{|\pi_j|} \sum_{a_i \in \pi_j} y_i \right). \tag{4}$$

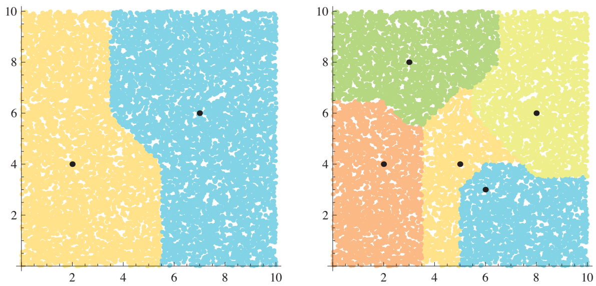
2.2 Izravni problem lokacije uz d_1 metričku funkciju

Za dani skup točaka $S \subset \mathbb{R}^2$ potražit ćemo rješenje izravnog problema lokacije *A* i izravnog problema *B* uz korištenje d_1 metričke funkcije. Ako su poznati centri $c_1,\ldots,c_k \in \mathbb{R}^2$, particiju $\Pi = \{\pi_1,\ldots,\pi_k\}$ određujemo *principom minimalnih udaljenosti* tako da bude

$$a_i \in \pi_j \Leftrightarrow \|c_j - a_i\|_1 \le \|c_s - a_i\|_1, \forall s = 1,\ldots,k. \tag{5}$$

Na Slici 2 ovaj problem ilustriran je primjenom vlastitog `Mathematica-modula` za $m = 10\,000$ podataka u ravnini i $k = 2$, odnosno $k = 5$ centara.

a) $m = 10\,000$, $k = 2$ b) $m = 10\,000$, $k = 5$



Slika 2: *Grupiranje skupa S u klastere*

Ako je poznata particija $\Pi = \{\pi_1, \ldots, \pi_k\}$ skupa $S = \{a_i = (x_i, y_i) \in \mathbb{R}^2 : i = 1, \ldots, m\} \subset \mathbb{R}^2$, centre klastera $c_1, \ldots, c_k$ dobivamo na sljedeći način (vidi [17])

$$c_j = \underset{z \in \mathrm{conv}(\pi_j)}{\mathrm{argmin}} \sum_{a_i \in \pi_j} \|z - a\|_1 = \left(\underset{a_i \in \pi_j}{\mathrm{med}}\, x_i,\ \underset{a_i \in \pi_j}{\mathrm{med}}\, y_i \right), \tag{6}$$

gdje je $\underset{a_i \in \pi_j}{\mathrm{med}}\, x_i$ oznaka za medijan niza kojeg čine apscise svih točaka iz klastera π_j.

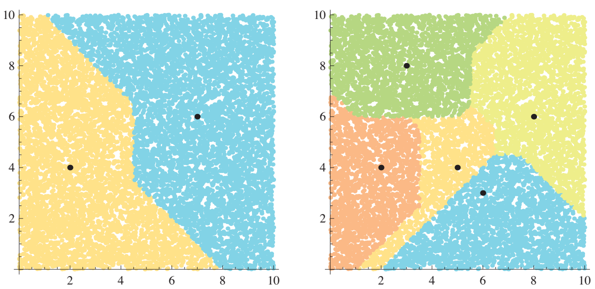
2.3 Izravni problem lokacije uz d_∞ metričku funkciju

Za dani skup točaka $S \subset \mathbb{R}^2$ potražit ćemo rješenje izravnog problema lokacije A i izravnog problema B uz korištenje d_∞ metričke funkcije. Ako su poznati centri $c_1, \ldots, c_k \in \mathbb{R}^2$, particiju $\Pi = \{\pi_1, \ldots, \pi_k\}$ određujemo *principom minimalnih udaljenosti* tako da bude

$$a_i \in \pi_j \Leftrightarrow \|c_j - a_i\|_\infty \le \|c_s - a_i\|_\infty, \quad \forall s = 1, \ldots, k. \tag{7}$$

Na Slici 3 ovaj problem ilustriran je primjenom vlastitog `Mathematica-modula` za $m = 10\,000$ podataka u ravnini i $k = 2$, odnosno $k = 5$ centara.

a) $m = 10\,000$, $k = 2$ b) $m = 10\,000$, $k = 5$



Slika 3: *Grupiranje skupa S u klastere*

Ako je poznata particija $\Pi = \{\pi_1, \ldots, \pi_k\}$ skupa $S = \{a_i = (x_i, y_i) \in \mathbb{R}^2 : i = 1, \ldots, m\} \subset \mathbb{R}^2$, centre klastera $c_1, \ldots, c_k$ dobivamo na sljedeći način

$$c_j = \underset{z \in \mathrm{conv}(\pi_j)}{\mathrm{argmin}} \sum_{a \in \pi_j} \|z - a\|_\infty. \tag{8}$$

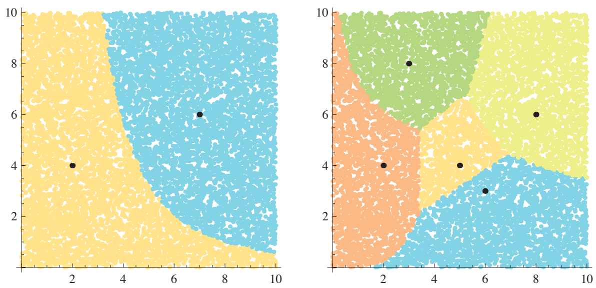
2.4 Izravni problem lokacije uz Bregmanovu generaliziranu I-udaljenost

Za dani skup točaka $S \subset \mathbb{R}^2$ potražit ćemo rješenje izravnog problema lokacije A i izravnog problema B uz korištenje kvazimetričke funkcije d_B koja je u literaturi poznata kao Bregmanova generalizirana I-udaljenost. Ako su poznati centri $c_1, \ldots, c_k \in \mathbb{R}^2$, particiju $\Pi = \{\pi_1, \ldots, \pi_k\}$ određujemo *principom minimalnih udaljenosti* tako da bude

$$a_i \in \pi_j \Leftrightarrow d_B(c_j, a_i) \le d_B(c_s, a_i), \forall s = 1, \ldots, k. \tag{9}$$

Na Slici 4 ovaj problem ilustriran je primjenom vlastitog `Mathematica-modula` za $m = 10\,000$ podataka u ravnini i $k = 2$, odnosno $k = 5$ centara. Primijetimo da su rubovi dobivenih disjunktnih skupova krivulje.

a) $m = 10\,000$, $k = 2$ b) $m = 10\,000$, $k = 5$



Slika 4: *Grupiranje skupa S u klastere*

Ako je poznata particija $\Pi = \{\pi_1, \ldots, \pi_k\}$ skupa $S = \{a_i = (x_i, y_i) \in \mathbb{R}^2 : i = 1, \ldots, m\} \subset \mathbb{R}^2$, centre klastera $c_1, \ldots, c_k$ dobivamo primjenom geometrijske sredine na podatke iz klastera

$$c_j = \underset{z \in \mathrm{conv}(\pi_j)}{\mathrm{argmin}} \sum_{a \in \pi_j} d_B(z, a) = \left(\left(\prod_{a_i \in \pi_j} x_i \right)^{1/|\pi_j|}, \left(\prod_{a_i \in \pi_j} y_i \right)^{1/|\pi_j|} \right). \tag{10}$$

2.5 `Mathematica-modul`

Sve prikazane ilustracije izrađene su vlastitim *Mathematica–modulom* `Particija[a, c, d]`. Modulu se predaje lista podataka `a`, lista centara `c` i ranije definirana kvazimetrička funkcija `d`. Za svaki element `a[[i]]` liste `a` modul pronalazi najbliži centar `c[[j]]`, a nakon toga element `a[[i]]` pridružuje klasteru `pi[[j]]`. Elementi svakog klastera prikazuju se točkicama jedne boje. Centri klastera označeni su crnim točkama.

```
In[1]:= Particija[a_, c_, d_] :=
Module[{m=Length[a], k=Length[c], pi,tab,min,imin},
 slc=ListPlot[c,
     PlotStyle -> {Black, AbsolutePointSize[5]}];
                    (* separacija *)
 pi = Table[{}, {j, k}];
 Do[
    tab = Table[d[c[[j]], a[[i]]], {j, k}];
    min = Min[tab];
    imin = Position[tab, min][[1]];
```

```
 Do[
    If[imin=={j}, pi[[j]]=Append[pi[[j]], a[[i]]]],
       {j, k}],
       {i, m}];
                 (* crtanje *)
 tab = Table[
        ListPlot[pi[[j]],
          PlotStyle->{Opacity[.5], Hue[.13 j^2]}],
           {j, k}];
Show[tab,slc, AspectRatio->Automatic, ImageSize->200]
     ]
```

Lista podataka `a` treba biti sastavljena od objekata oblika $\{x,y\}$. To znači da podaci mogu dolaziti iz proizvoljno odabranog područja $\Omega \subset \mathbb{R}^2$. Primjerice, naredbom

```
RandomReal[10, {10000, 2}]
```

definira se 10 000 uniformno distribuiranih slučajnih točaka iz kvadrata $[0,10] \times [0,10]$. Lista centara `c` odabire se po volji. Kvazimetrička funkcija `d` definira se kao funkcija dviju varijabli. Primjerice, kvazimetrička funkcija najmanjih kvadrata d_{LS} definira se naredbom

```
d[x_, y_] := Norm[x - y, 2]^2
```

a metrička funkcija d_1 naredbom

```
d[x_, y_] := Norm[x - y, 1]
```

Podatke, kvazimetričku funkciju i poziv modula, primjerice za slučaj prikazan na Slici 2b, možemo implementirati na sljedeći način.

```
In[2]:= SeedRandom[3]
        a = RandomReal[10, {10000, 2}];
        c = {{5,4}, {6,3}, {8,6}, {2,4}, {3,8}};
        d[x_, y_] := Norm[x - y, 1]
        Particija[a, c, d]
```

Pri tome naredba `SeedRandom[3]` osigurava da ćemo svakim pokretanjem programa dobiti iste slučajne brojeve. Izvođenje programa za navedene primjere traje $1-2$ sekunde.

3 Obratni problem lokacije

Zadan je skup točaka $S = \{a_i = (x_i, y_i) \in \mathbb{R}^2 \colon i = 1,\ldots,m\} \subset \mathbb{R}^2$ u ravnini i neka kvazimetrička funkcija $d \colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}_+$. Treba pronaći optimalne centre $c_1^\star,\ldots,c_k^\star \in \mathbb{R}^2$ tako da bude

$$F(c_1^\star,\ldots,c_k^\star) = \min_{c_1,\ldots,c_k \in \mathrm{conv}(S)} F(c_1,\ldots,c_k), \tag{11}$$

gdje je $F \colon \mathbb{R}^{2k} \to \mathbb{R}$

$$F(c_1,\ldots,c_k) = \sum_{i=1}^{m} \min_{1\le j\le k} d(c_j,a_i). \tag{12}$$

Poznavajući centre $c_1^\star,\ldots,c_k^\star$ principom minimalnih udaljenosti moguće je definirati optimalnu particiju $\Pi^\star = \{\pi_1^\star,\ldots,\pi_k^\star\}$ (problem A u *t.2*). Primjerice problem određivanja optimalnih izbornih jedinica iz Primjera 4 jedan je obrnuti problem lokacije, a može se definirati na sljedeći način.

$$F(c_1,\ldots,c_k) = \sum_{i=1}^{n} w_i \min_{1\le j\le k} d(c_j,a_i) \quad \longrightarrow \min,$$

uz uvjete:

(i) $\quad (1-p)\frac{m}{k} \le |\pi_j| \le (1+p)\frac{m}{k}, \quad j = 1,\ldots,k$

(ii) $\quad d(c_j,a_i) \le r, \quad \forall a_i \in \pi_j, \forall j = 1,\ldots,k.$

Kao što smo već ranije spomenuli, ovaj problem u literaturi se može pronaći pod nazivom problem k-sredina ili problem k-medijana, a u općem slučaju radi se o problemu traženja globalnog minimuma višedimenzionalne nediferencijabilne funkcije (vidi [5, 6]) koja može imati veći broj lokalnih minimuma (vidi primjerice [11, 15]). O metodama za rješavanje ovog problema može se vidjeti primjerice u [16, 18].

Najpoznatiji postupak za rješavanje ovog problema je *algoritam k-sredina* (vidi [7, 11, 13, 15, 17, 19, 20]), kojim nažalost možemo pronaći lokalni minimum kriterijske funkcije cilja. Ipak, višestrukim pokretanjem ovog algoritma s različitim početnim aproksimacijama, možemo pronaći optimalno rješenje (vidi [15]). Niže navodimo skicu algoritma k-sredina za opći slučaj kada je $S \subset \mathbb{R}^n$.

Algoritam 1.
(Standardni algoritam k-sredina)[3]

Korak 1: Izabrati $\Pi = \{\pi_1,\ldots,\pi_k\}$;

Korak 2: Izračunati: $\theta = (c_1,\ldots,c_k)$, pri čemu je $c_j = \underset{z\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{a_i\in\pi_j} d(z,a_i)$;
Izračunati $F_1 = F(\theta)$;

Korak 3: Pomoću centara iz θ prema principu minimalnih udaljenosti formirati novu particiju $\mathcal{N} = \{\nu_1,\ldots,\nu_k\}$;
Izračunati nove centre: $\zeta = (\zeta_1,\ldots,\zeta_k)$, pri čemu je $\zeta_j = \underset{z\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{a_i\in\nu_j} d(z,a_i)$;
Izračunati $F_2 = F(\zeta)$;

Korak 4: Ako je $F_2 < F_1$, staviti $\theta = \zeta$; $F_1 = F_2$ i prijeći na Korak 3; u suprotnom STOP.

[3]Odgovarajuća programska podrška dostupna je na adresi: `http://www.mathos.hr/oml/software.htm`

Literatura

[1] M. R. Ackermann, J. Blömer, *Coresets and Approximate Clustering for Bregman Divergences*, Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'09), 1088-1097, Society for Industrial and Applied Mathematics (SIAM), 2009

[2] A. Banerjee, S. Merugu, I. S. Dhillon, J. Ghosh, *Clustering with Bregman divergences*, Journal of Machine Learning Research, **6**(2005), 1705–1749

[3] P. Chardaire, *Facility location optimization and cooperative games*, PhD thesis, University of East Anglia, Norwich, UK, 1998.

[4] B. Divjak, *Bilješke o taxicab geometriji*, KoG **5**(2000), 5-9

[5] D. E. Finkel, C. T. Kelley, *Additive scaling and the DIRECT algorithm*, J. Glob. Optim. **36**(2006), 597-–608

[6] C. A. Floudas, C. E. Gounaris, *A review of recent advances in global optimization*, J. Glob. Optim. **45**(2009), 3-–38

[7] G. Gan, C Ma, J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, SIAM, Philadelphia, 2007.

[8] Ž. Gjuranić, *Modeliranje terena pomoću Delaunayjeve triangulacije*, KoG **11**(2007), 49-52

[9] M. X. Goemans, M. Skutella, *Cooperative facility location games*, Journal of Algorithms **50**(2004) 194–214

[10] K. Graham, D. E. Knuth, O. Patashnik, *Concrete Mathematics*, Addison - Wesley, Boston, 2003.

[11] C. Iyigun, *Probabilistic Distance Clustering*, Dissertation, Graduate School – New Jersey, Rutgers, 2007.

[12] C. Iyigun, A. Ben-Israel, *A generalized Weiszfeld method for the multi-facility location problem*, Operations Research Letters **38**(2010), 207-214

[13] J. Kogan, *Introduction to Clustering Large and High-Dimensional Data*, Cambridge University Press, 2007.

[14] J. Kogan, C. Nicholas, M. Wiacek, *Hybrid clustering with divergences*. In: M. W. Berry and M. Castellanos (Eds.), Survey of Text Mining: Clustering, Classification, and Retrieval, Second Edition, Springer, 2007.

[15] F. Leisch, *A toolbox for K-centroids cluster analysis*, Computational Statistics & Data Analysis **51**(2006), 526-544

[16] A. M. Rodrígues-Chia, I. Espejo, Z. Drezner, *On solving the planar k-centrum problem with Euclidean distances*, European Journal of Operational Research **207**(2010), 1169-1186

[17] K. Sabo, R. Scitovski, I. Vazler, *Grupiranje podataka: klasteri*, Osječki matematički list **10**(2010), 149-176

[18] A. Schöbel, D. Scholz, *The big cube small cube solution method for multidimensional facility location problems*, Computers & Operations Research **37**(2010), 115–122

[19] H. Späth, *Cluster–Formation und- Analyse*, R. Oldenburg Verlag, München, 1983.

[20] M. Teboulle, *A unified continuous optimization framework for center-based clustering methods*, Journal of Machine Learning Research **8**(2007), 65-102

[21] S. Toprak, E. Nacaroglu, O. A. Cetin, A. C. Koc, *Pipeline Damage Assessment Using Cluster Analysis*, TCLEE 2009: Lifeline Earthquake Engineering in a Multihazard Environment, pp. 1-8, (doi 10.1061/41050(357)78)

[22] D. Veljan, *Kombinatorna i diskretna matematika*, Algoritam, Zagreb, 2001.

Kristian Sabo

Odjel za matematiku
Sveučilište J. J. Strossmayera u Osijeku
e-mail: ksabo@mathos.hr

Sanja Scitovski

Risnjačka 7, Osijek
e-mail: sanja@mathos.hr

Zahvala: Zahvaljujemo anonimnom recenzentu, koji je svojim primjedbama i sugestijama u znatnoj mjeri pomogao da ovaj tekst bude bolji.