



KLASTEROVANJE KADA PODACI NEDOSTAJU KORIŠĆENJEM METODE PROMENLJIVIH OKOLINA

CLUSTERING WHEN MISSING DATA BY USING THE VARIABLE NEIGHBORHOOD SEARCH

NATAŠA GLIŠOVIĆ¹, TATJANA DAVIDOVIĆ², MIODRAG RAŠKOVIĆ²

¹ Državni Univerzitet u Novom Pazaru, natasaglisovic@gmail.com

² Matematički Institut SANU, [{tanjad, goca}@mi.sanu.ac.rs}](mailto:{tanjad, goca}@mi.sanu.ac.rs)

Rezime: Rad se bavi istraživanjem efikasnosti klasterovanja pomoću metode promenljivih okolina (eng. *Variable neighborhood search-VNS*), konkretno, pomoću nekih njenih varijanti. Takođe, bavi se i problemom nedostajućih podataka i donošenja odluke u takvim okolnostima. Cilj ovog istraživanja je da se predloženo rastojanje, u slučajevima kada podaci iz nekog razloga nedostaju, implementira kroz neke varijante metaheurističke metode promenljivih okolina. Efikasnost implementacije pokazana je kroz eksperimentalne rezultate na konkretnoj bazi autoimunih bolesti Kliničkog centra Srbije.

Ključne reči: metod promenljivih okolina (VNS), redukovana metoda promenljivih okolina (RVNS), nedostajući podaci, rastojanje, klasterovanje, autoimune bolesti.

Abstract: The paper deals with the study of the clustering efficiency with the variable neighborhood search (VNS), more precisely with some of its variants. It also deals with the problem of missing data and decision making in such circumstances. The aim of this research is to implement the proposed distance, in cases when data for some reason are missing, into some variants of the variable neighborhood search metaheuristic method. The efficiency of the proposed implementation is demonstrated through the experimental results on the concrete basis of autoimmune diseases from the Clinical Center of Serbia..

Keywords: Variable neighborhood search (VNS), Reduced Variable neighborhood search (RVNS), Missing Data, Distance, Clustering, Autoimmune Diseases.

1. UVOD

Mogućnosti primene operacionih istraživanja u medicini su mnogostruke, na primer postavljanje dijagnoza, doziranje terapije, optimizacija dužine lečenja i slično. U ovom radu razmatra se primena optimizacionih metoda u postavljanju dijagnoze bolesnicima na osnovu izvršenih pregleda i analiza. Svaki pacijent predstavljen je vektorom atributa koji se sastoji od rezultata izvršenih analiza. Dijagnoza se ustvari svodi na klasterovanje pacijenata u zadati broj bolesti na osnovu rastojanja definisanih nad odgovarajućim vektorima.

Klasterovanje (eng. clustering) je tehnika istraživanja podataka koja otkriva objekte (koji se opisuju atributima) sličnih osobina i deli ih u grupe (klaster), čineći ih preglednijim i korisnijim. Klaster analiza zapravo predstavlja pronalaženje grupa objekata takvih da su objekti u grupi međusobno slični (ili povezani), a da su objekti u različitim grupama međusobno raličiti (ili nepovezani).

Metode zasnovane na udaljenosti su veoma popularne u literaturi, jer mogu da se koriste za bilo koji tip podataka, dok god postoji odgovarajuća funkcija rastojanja pogodna za tu vrstu podataka. Dakle, problem grupisanja podataka može da se svede na problem pronalaženja funkcije rastojanja za tu vrstu podataka. Iz ovoga proizilazi da je pronalaženje funkcije rastojanja postala važna oblast istraživanja u obradi podataka (Wang and Sun 2012; Das and Mannila 2000).

Da bi se uopšte sprovela analiza grupisanja, neophodno je definisati mere bliskosti dva objekta na osnovu njihovih karakteristika. Koncept „sličnosti” se određuje u zavisnosti od samih podataka. S obzirom da su podaci u većini slučajeva vektori stvarnih vrednosti, Euklidska udaljenost između podataka može poslužiti kao mera te različitosti. Međutim, jedan od nedostataka poznatih rastojanja (kao što su: Euklidsko rastojanje, Menhetn rastojanje, rastojanje Minkovskog) je što se mogu primeniti samo kada nam je poznata vrednost svih komponenti, tj. osobina koje posmatramo kod objekata (Aggarwal 2003).

Osnovni problem pri primeni standardnih metoda klasterovanja na problem razmatran u ovom radu je što u medicini nije lako odrediti rastojanje s obzirom da veoma često neke od analiza nedostaju. Zato se koristi nova funkcija merenja rastojanja predložena u prethodnim istraživanjima (Glišović and Rašković 2017) koja je ugrađena u model celobrojnog linearnog programiranja (eng. Integer Linear Programming, ILP) i nove heurističke metode zasnovane na metaheurističkoj metodi promenljivih okolina. Implementirane metode primenjene su u klasterovanju medicinskih podataka Kliničkog centra Srbije, odeljenja za alergologiju i imunologiju kod kojih se javio problem nedostajućih podataka.

Istraživanje je prikazano kroz nekoliko faza. Pošto baza podataka koja je korišćena u eksperimentalnom delu ima nedostajuće podatke za pojedine attribute (koji opisuju svaki element baze, objekat), u ovom istraživanju smo prvo opisali predloženo rastojanje na osnovu kojeg možemo odrediti udaljenost dva objekta kada podaci nedostaju. U narednom odeljku će biti reči o problemu nedostajućih podataka kao i izboru mera sličnosti/rastojanja. Zatim ćemo formulisati problem klasterovanja kao i primenu osnovne i redukovane metode promenljivih okolina na problem klasterovanja sa nedostajućim podacima. Rezultati istraživanja biće predstavljeni u odeljku 4, a zaključak i mogućnosti za neka dalja istraživanja predstavljemo u odeljku 5.

2. PROBLEM NEDOSTAJUĆIH PODATAKA. IZBOR MERA SLIČNOSTI/RASTOJANJA

Problem nedostajućih podataka je od velikog značaja. Kada se govori o mehanizmima nedostajućih podataka treba utvrditi prvo razlog nedostajanja podataka. Podaci mogu nedostajati iz više razloga. Neki od njih su: podaci nisu raspoloživi, došlo je do grešaka u radu sa opremom, nekonzistentnosti sa drugim podacima, pa su zato izbrisani, nisu unešeni zbog nerazumevanja, nisu smatrani bitnim u trenutku unosa itd. Bitna je odluka šta raditi sa nedostajućim podacima. Neke od mogućnosti su (Graham 2012):

- Izbrisati elemente kod kojih se javljaju nedostajući podaci-što nije preporučljivo posebno kod klasifikacije, a naročito ako nedostajuće vrednosti variraju od elementa do elementa, tj. nedostaju različiti elementi kod različitih objekata (vektora).
- Ručno popunjavanje nedostajućih vrednosti koje je zamorno i često neizvodljivo.
- Automatsko popunjavanje: nekom opštom konstantom, srednjom vrednosti elemenata za sve objekte (vektore) koji pripadaju istoj klasi.
- Najverovatnija vrednost-zaključak se donosi na osnovu Bajesove formule ili prema stablu odlučivanja.

U bazi sa kojom se radilo u ovom istraživanju (baza pacijenata koji boluju od autoimunih bolesti) do nedostajućih podataka je došlo zato što za neke od pacijenata nije bilo potrebe da se odrade određene analize, jer se pre njih ustanovilo o kojoj bolesti je reč ili nije bilo sredstava zato što su pojedine analize skupe, a neki od podataka nedostaju zbog gubitka tog podatka u trenutku unosa u bazu podataka (u otpusnim listama na osnovu kojih su unošeni podaci došlo je do previda).

Da bi prevazišli ovo ograničenje, a da opet sačuvamo kompaktnost nalaza tj. da ne vršimo brisanje nalaza kod kojih postoje nedostajući podaci, predlažemo da se koristi rastojanje definisano formulom (1) predloženo u Glišović and Rašković (2017).

Da bi se uopšte sproveda analiza grupisanja, neophodno je definisati mere bliskosti dva objekta na osnovu njihovih karakteristika. Koncept „sličnosti” se određuje u zavisnosti od samih podataka. S obzirom da su podaci u većini slučajeva vektori stvarnih vrednosti, Euklidska udaljenost između podataka može poslužiti kao mera te različitosti.

Neka baza podataka sadrži m objekata $x_i \in R^n$, $i = 1, \dots, m$. Za meru $d : R^n \times R^n \rightarrow R$ kažemo da predstavlja *meru različitosti* (rastojanja, metriku), objekata $x_i \in R^n$ i $y_j \in R^n$ u oznaci $d_{ij} = d(x_i, y_j)$ ako zadovoljava sledeće osobine za sve objekte:

1. $d_{ij} \geq 0$ (uslov nenegativnosti)
2. $d_{ij} = d_{ji}$ (uslov simetričnosti)
3. $d_{ij} \leq d_{ik} + d_{kj}$ (uslov nejednakosti trougla)

Za meru $s_{ij} = s(x_i, y_j)$ kažemo da predstavlja *meru sličnosti* objekata x_i i y_j , gde je $1 \leq i, j \leq m$, ako zadovoljava sledeće osobine:

1. $0 \leq s_{ij} \leq 1$ (uslov normiranosti)
2. $s_{ij} = 1$, samo ako su objekti jednaki

3. $s_{ij} = s_{ji}$ (uslov simetričnosti)

Pored problema nedostajućih podataka opisanog iznad, unešeni podaci u bazi koju smo koristili su šifrovani, pa se zato došlo na ideju korišćenja Hamingovog rastojanja (eng. Hamming distance) (Norouzi *et al* 2012) i predloga novog rastojanja koje koristi Hamingovo.

Neka je F konačan skup vektora.

Definicija: Hamingovo rastojanje $d(\underline{x}, \underline{y})$ između dva vektora $\underline{x}, \underline{y} \in F^{(n)}$ je broj mesta (elemenata) na kojim se ova dva vektora razlikuju.

Rastojanje predloženo u Glišović and Rašković (2017) koristi Hamingovo rastojanje i iskazne formule. Korišćenje iskaznih formula u definiciji rastojanja je zbog opštosti rastojanja.

Neka su φ i ψ dva skupa iskaznih formula kojem pripadaju formule koje predstavljaju konjunkciju literala. Predloženo rastojanje između ova dva skupa iskaznih formula definisano je sa:

$$D(\varphi, \psi) = \frac{\max_{A \in \varphi} \min_{B \in \psi} d(A, B) + \max_{B \in \psi} \min_{A \in \varphi} d(A, B)}{2} \quad (1)$$

gde je $d(A, B)$ Hamingovo rastojanje.

Primer: Neka su data dva vektora v_1 i v_2 dužine tri sa vrednostima elemenata a , b i c na takav način da imamo sledeće podatke o vektorima:

vektor v_1 ima $a \wedge b$ (iz nekog razloga nedostaje informacija o vrednosti elementa c)

vektor v_2 ima $a \wedge \neg b \wedge c$

Ako primenimo predloženo rastojanje, formula φ koja opisuje prvi vektor, je: $a \wedge b \wedge (c \vee \neg c)$.

Kada pretvorimo sve u konjukcije imećemo dve formule $a \wedge b \wedge c$ i $a \wedge b \wedge \neg c$ (te dve formule pripadaju skupu formula φ). Dok skup formula drugog vektora, skup ψ , čini samo formula $a \wedge \neg b \wedge c$. Ako primenimo rastojanje (1), na ova dva vektora, dobićemo da ono iznosi:

$$\begin{aligned} D(\varphi, \psi) &= \frac{\max_{A \in \varphi} \min_{B \in \psi} d(A, B) + \max_{B \in \psi} \min_{A \in \varphi} d(A, B)}{2} = \\ &= \frac{\max \left\{ \min \{d(a \wedge b \wedge c, a \wedge \neg b \wedge c)\}, \min \{d(a \wedge b \wedge \neg c, a \wedge \neg b \wedge c)\} \right\} + \max \left\{ \min \{d(a \wedge \neg b \wedge c, a \wedge b \wedge c), d(a \wedge \neg b \wedge c, a \wedge b \wedge \neg c)\} \right\}}{2} \\ &= \frac{\max \left\{ \min \{1\}, \min \{2\} \right\} + \max \left\{ \min \{1, 2\} \right\}}{2} = \frac{2+1}{2} = \frac{3}{2}. \end{aligned}$$

3. FORMULACIJA PROBLEMA KLSTEROVANJA. OSNOVNA I REDUKOVANA METODA PROMENLJIVIH OKOLINA

3.1. KLASTEROVANJE

Postoji nekoliko formulacija problema klasterovanja zavisno od funkcije cilja koja se optimizuje, a u ovom istraživalju je korišćena formulacija preko problema p-medijane (Mladenović *et al.* 2007).

Neka su x_{ij} i y_j binarne promenljive definisane na sledeći način:

$$\begin{aligned} x_{ij} &= \begin{cases} 1, & \text{ako se objekat } i \text{ nalazi u klasteru } j, \\ 0, & \text{inače.} \end{cases} \\ y_j &= \begin{cases} 1, & \text{ako objekat } j \text{ reprezentuje odgovarajući klaster,} \\ 0, & \text{inače.} \end{cases} \\ \min \sum_i \sum_j d_{ij} x_{ij} & \quad (2) \end{aligned}$$

t. d.

$$\sum_j x_{ij} = 1 \text{ za svako } i \quad (3)$$

$$x_{ij} \leq y_j \text{ za svako } i, j \quad (4)$$

$$\sum_j y_j = p \quad (5)$$

$$x_{ij}, y_j \in \{0,1\} \quad (6)$$

Formulacija je iskorišćena u komercijalnom solveru CPLEX za dobijanje optimalnog rešenja kao referentnog rezultata za utvrđivanje performanse VNS metoda.

Problem koji se u ovom radu razmatra odnosi se na klasterovanje pacijenata u cilju postavljanja odgovarajućih dijagnoza. Svaki pacijent opisan je vektorom nalaza među kojima neki od elemenata mogu nedostajati. Broj klastera (mogućih bolesti) zadat je unapred.

U bazi sa kojom se radilo u ovom istraživanju (baza pacijenata koji boluju od autoimunih bolesti) ima 45 pacijenata koji su opisani kroz 89 nalaza. Primer nekih od nalaza sa kodiranim odgovarajućim vrednostima dat je u Tabeli 1. Baza je kodirana i ima nedostajuće podatke tako da smo u postupku klasterovanja primenili rastojanje (1) opisano u odeljku 2. Ukoliko neki nalaz nedostaje, odgovarajuća vrednost u vektoru ne postoji.

Tabela 1: Nazivi i vrednosti nekih od 89 nalaza kao ilustrativni primer kodirane baze.

Nalaz	Vrednost
Coombsov test	0. negativan; 1. Pozitivan
ANA	0. negativan; 1. pozitivan<1:80; 2. pozitivan >1:80
PH biopsije bubrega	0. uredno; 1. mezancijalni N II; 2. fokalno prolif. III; 3. difuzno prolif. IV; 4.membranski GN V; 5. skleroza (UZ)

Pacijenti se svrstavaju u tri grupe (klastera) jer se u bazi nalaze pacijenti koji boluju od tri bolesti-sistemska skleroza, sjegren i lupus. Svaki klaster karakteriše se svojim centroidom koji je u našem slučaju realni pacijent. To znači da se pacijenti grupišu na osnovu rastojanja od izabranog centroida.

3.2. OSNOVNA I REDUKOVANA METODA PROMENLJIVIH OKOLINA

Metoda promenljivih okolina je metaheuristika koja je predstavljena devedesetih godina prošlog veka (Mladenović 1995, Mladenović and Hansen 1997) nakon čega je doživela mnogo promena i ekstenzija (Hansen *et al.* 2010, Hansen *et al.* 2010a), kao i uspešnih primena (Hansen and Mladenović 2012). VNS metaheuristika zasnovana je na tri osnovne činjenice:

1. Lokalni minimum u odnosu na jednu okolinu ne mora biti i lokalni minimum u odnosu na neku drugu okolinu;
2. Globalni minimum je lokalni minimum u odnosu na sve okoline;
3. Za većinu problema lokalni minimumi u odnosu na razne okoline su međusobno bliski.

Najjednostavnija varijanta VNS metode je redukovana metoda promenljivih okolina (RVNS). Ona se sastoji u sistemskoj promeni okolina i izboru jednog slučajnog rešenja u svakoj od okolina. Koraci odlučivanja bazirani su na tom jednom slučajnom rešenju. Ova metoda je izuzetno korisna za dobijanje početnih rešenja kod primera velikih dimenzija i najbolji rezultat se dobija njenom kombinacijom sa nekom drugom varijantom. Pesudokod RVNS metode dat je na slici 1.

Uobičajeni kriterijum zaustavljanja je maksimalni broj iteracija između dva poboljšanja.

Osnovna metoda promenljivih okolina (BVNS) je najrasprostranjenija varijanta metode promenljivih okolina jer obezbeđuje više preduslova za dobijanje kvalitetnijih konačnih rešenja. Kod BVNS metode osnovni koraci sadržani su u petlji u kojoj menjamo indeks okoline i , određujemo slučajno rešenje iz te okoline, izvršavamo proceduru lokalnog pretraživanja počev od tog slučajnog rešenja i proveravamo kvalitet dobijenog lokalnog minimuma. Ove korake ponavljamo dok ne bude zadovoljen neki od kriterijuma zaustavljanja. Prilikom svakog odabira okoline početna rešenja generišemo na slučajan način kako bi obezbedili pretraživanje različitih regiona kod svakog sledećeg razmrđavanja u okolini i . Okoline razlikujemo po broju transformacija (rastojanju) ili po vrsti transformacija (metrici).

Napominjemo da okoline za izbor slučajnog rešenja (razmrđavanje) i lokalno pretraživanje ne moraju biti istog tipa. Pseudokod BVNS metode dat je na slici 2.

RVNS i BVNS primenjeni na problem klasterovanja pacijenata implementirani su na sledeći način. Najpre su izračunata rastojanja između pacijenata primenom formule (1). U fazi predprocesiranja vrste matrice rastojanja (kao i odgovarajući indeksi pacijenata) sortirane su u neopadajućem poretku. Ovi podaci korišćeni su za efikasniju implementaciju operatora razmrđavanja. Naime, kako se svako rešenje karakteriše skupom centroida, operator razmrđavanja sastoji se u zameni odgovarajućeg broja centroida. Preciznije, razmrđavanje u okolini k podrazumeva da se centroidi zamene slučajno izabranim pacijentima koji nisu centroidi, a udaljeni su najviše k mesta od centroida koga menjaju. U svakom koraku razmatra se zamena

svih centroida, pri čemu do zamene neće doći ukoliko je slučajno izabrani pacijent najbliži tom centroidu (ustvari to je sam taj centroid). Lokalno pretraživanje sastoji se u sistematskoj zameni jednog centroida pacijentom koji nije centroid. Ono polazi od rešenja dobijenog razmrdavanjem i izvršava se po *best improvement* principu dok god ima poboljšanja.

Inicijalizacija. Izabrati početno rešenje X , $x_{optimalno} = X$, $f_{optimalno} = f(X)$.

Radi

```
{
    i = 1
    Radi
    {
        Izabrati slučajno rešenje  $x'$  u okolini rešenja  $X$  (Razmrdavanje(i))
        Ako je  $f(x') < f(x_{optimalno})$  onda je
             $x_{optimalno} = x'$ 
             $f(x_{optimalno}) = f(x')$ 
            i = 1
        inače i = i + 1
    } sve dok nije i = imax
} sve dok nije zadovoljen kriterijum zaustavljanja
```

Slika 1. Pseudokod RVNS metode

Inicijalizacija. Izabrati početno rešenje $x \in X$ i definisati kriterijum zaustavljanja STOP=0

Ponavljaj

```
{
    i = 1
    Ponavljaj
    {
        Generisati slučajno rešenje  $x'$  u  $i$ -toj okolini od  $X$ - Razmrdavanje(i).
         $x'' = \text{Lokalno pretraživanje}(x')$ 
        Ako je lokalni minimum bolji od trenutnog minimuma
             $x_{optimalno} = x''$ 
             $f(x_{optimalno}) = f(x'')$ 
            i = 1
        Inače preći u i = i + 1
        Ako je zadovoljen kriterijum zaustavljanja
            STOP=1.
    } dok nije i = imax ili STOP=1
} sve dok nije STOP=1
```

Slika 2. Pseudokod BVNS metode

4. REZULTATI ISTRAŽIVANJA

RVNS i BVNS metode implementirane su u C# programskom jeziku na računaru HP-15-d055, pod operativnim sistemom Windows 10 Pro. S obzirom na stohastičku prirodu metoda vršeno je 100 restartovanja, a kao kriterijum zaustavljanja zadato je vreme izvršavanja CPLEX komercijalnog solvera (1.09 s). Optimalno rešenje CPLEX-a je 82.19, što je i postignuto i kod obe varijante VNS-a, ali za znatno kraće vreme (Tabela 2). Napominjemo da su vremena data u tabeli 2 izražena u milisekundama.

Implementirane VNS metode pokazale su veliku stabilnost jer je od 100 puštanja u BVNS-u najbolje rešenje dostignuto 86 puta, dok je kod RVNS-a to bilo 83 puta. Na osnovu rezultata prikazanih u tabeli 2 može se zaključiti da je BVNS stabilnija varijanta dok je RVNS značajnije brža uz neznatan gubitak na stabilnosti.

Tabela 2: Rezultati RVNS-a i BVNS-a.

Metoda	Vrednost najboljeg rešenja	Prosečna vrednost funkcije cilja	Prosečno vreme (ms)
RVNS	82.19	86.66	0.13
BVNS	82.19	82.24	3.36

Ako se analizira uspešnost svrstavanja pacijenata u odgovarajuće klastere upoređena sa originalnom bazom dobije se da je kod RVNS-a i BVNS-a 91.1% pacijenata ispravno određena bolest (od 45 pacijenata 4 su raspoređena u klastere kojima ne pripadaju). Medicinski gledano, ta četiri bolesnika su bila iz grupe lupusa (u izvornoj bazi sa tom dijagnozom), a raspoređeni su u sjegren klaster. Mada, uz konsultovanje stručnjaka iz medicine ove dve bolesti jesu bliske jedna drugoj i dešava se u praksi da pacijenti sa jednom dijagnozom „pređu“ vremenom u drugu (iz lupusa u sjegren i obrnuto).

5. ZAKLJUČAK

U ovom istraživanju cilj je bio da se predlog kalsterovanja sa nedostajućim podacima, tj. da se ovaj NP težak problem implementira kroz metaheurističku metodu promenljivih okolina (VNS) i pokaže njena uspešnost u rešavanju ovog problema na konkretnom problemu klasterovanja pacijenata koji boluju od autoimunih bolesti. Rezultati prikazani u tabeli 2 ukazuju da su metode zasnovane na VNS metaheuristici dostigle optimalno rešenje (dobijeno pomoću CPLEX solvera) za značajno kraće vreme.

Dalje istraživanje treba da obuhvati primene ovih metoda na veće baze podataka i njihovo poređenje sa drugim metodama. Takođe, istraživanje otvara mogućnost uključivanja težina svakom od nalaza i ispitivanje njihovih značaja u cilju povećavanja uspešnosti dobijenih rezultata u odnosu na ekspertski postavljene dijagnoze lekara specijalista.

LITERATURA

- [1] Das, G., & Mannila, H. (2000). Context-based similarity measures for categorical databases. PKDD Conference, 201–210.
- [2] Davidović, T. (2006). Raspoređivanje zadataka na višeprosesorske sisteme primenom metaheuristika [doktorska disertacija]. [Beograd]: Matematički fakultet
- [3] Glišović, N. and Rašković, M. (2017). Optimization for Classifying the Patients Using the Logic Measures for Missing Data, Scientific publications of the State University of Novi Pazar Ser. a: Appl. Math. Inform. and Mech. vol. 9, 1, 91-101.
- [4] Graham, J. W. (2012). Missing Data: Analysis and Design. Springer Science and Business Media, New York.
- [5] Hansen, P. and Mladenović, N. (2014). Variable neighborhood search. In Search methodologies (pp. 313-337). Springer US.
- [6] Hansen, P., Mladenović, N., Brimberg, J. and Perez, J. A. M. (2010). Variable neighborhood search Handbook of Metaheuristics ser. International Series in Operations Research & Management Science, 146, 61-86.
- [7] Hansen, P., Mladenović, N. and Pérez, J. A. M. (2010). Variable neighborhood search: methods and applications. Annals of Operations Research, 175(1), 367-407.
- [8] Mladenović, N. (1995). A Variable neighborhood algorithm – a new metaheuristic for combinatorial optimization, Abstracts of papers presented at Optimization Days, Montreal, p. 112.
- [9] Mladenović, N., Brimberg, J., Hansen, P., Moreno-Perez JA (2007). The p-median problem: a survey of metaheuristic approaches. European Journal of Operational Research 179:927–939
- [10] Mladenović, N., Hansen, P. (1997), Variable neighborhood search. Computers and Operations Research; 24(11); 1097–1100.
- [11] Norouzi, M., Fleet, D. J. and Salakhutdinov, R. R. (2012). Hamming distance metric learning. In Advances in neural information processing systems (pp. 1061-1069).
- [12] Wang, F. and Sun, J. (2012). Distance metric learning in data mining, SDM Conference (Tutorial)