

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

**Metodologija za usporedbu
kontekstualiziranih
polazišno-odredišnih matrica**

Rijeka, veljača 2020.

Vjera Turk
0069064924

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

**Metodologija za usporedbu
kontekstualiziranih
polazišno-odredišnih matrica**

Mentor: prof.dr.sc. Renato Filjar

Rijeka, veljača 2020.

Vjera Turk
0069064924

Rijeka, 17. ožujka 2019.

Zavod: **Zavod za računarstvo**
Predmet: **Usluge zasnovane na lokaciji**

ZADATAK ZA DIPLOMSKI RAD

Pristupnik: **Vjera Turk (0069064924)**
Studij: **Diplomski sveučilišni studij računarstva**
Modul: **Programsko inženjerstvo**

Zadatak: **Metodologija za usporedbu kontekstualiziranih polazišno-odredišnih matrica/Methodology for contextualised origin-destination matrices comparison**

Opis zadatka:

Polazišno-odredišna matrica, POM, omogućuje sustavnu statističku procjenu migracija stanovništva u zadanom prostorno-vremenskom okviru. Za razliku od tradicionalnog pristupa brojanja putovanja i putnika, za procjenu POM-e danas se sve više koristi statistička analiza podataka iz suvremenih informacijskih i komunikacijskih sustava (zapisi o aktivnostima u javnoj pokretnoj mreži, združena očitavanja prijamnika za satelitsku navigaciju i sl.), čime je omogućeno poboljšanje kvalitete procjene preslikavanjem POM-e na kontekst. Pojavljuje se potreba za objektivnom procjenom kvalitete POM-e u odnosu na referentnu (kontrolnu). U ovom radu potrebno je definirati odnosne parametre kvalitete POM-e te razviti metodologiju usporedbe dviju POM-a dobivenih različitim postupcima procjene i s podacima iz različitih izvora. Usporedbu obaviti korištenjem numeričkog i grafičkog oblika POM-e. Metodologiju je potrebno izvesti u programskom okruženju za statističko računarstvo R te demonstrirati njenu primjenu na slučaju usporedbe dviju POM-a. Komentirati dobivene rezultate sa stajališta apsolutne i relativne točnosti matrica.

Diplomski rad mora biti napisan prema Uputama za pisanje diplomskih radova koje su objavljene na mrežnim stranicama studija.

Polje znanstvenog područja: Računarstvo

Grana znanstvenog područja: Informacijski sustavi

Zadatak uručen pristupniku: 18. ožujka 2019.

Rok za predaju rada: 19. svibnja 2020.

Mentor:



red. prof. dr. sc. Renato Filjar

Predsjednik povjerenstva za
diplomski ispit:



izv. prof. dr. sc. Kristijan Lenac

Izjava o samostalnoj izradi rada

Izjavljujem da sam samostalno izradila ovaj rad.

Rijeka, veljača 2020.

Ime Prezime

Zahvala

Zahvaljujem mentoru prof. dr. sc. Renatu Filjaru na podršci tijekom pisanja ovoga rada, usmjeravanju i savjetima. Zahvaljujem obitelji i prijateljima na podršci tijekom studiranja i obitelji na razumijevanju i strpljenju tijekom pisanja ovog rada.

Sadržaj

Popis slika

Popis tablica

Poglavlje 1

Uvod

je alat koji omogućuje opis i sustavnu statističku procjenu migracija stanovništva na nekom području u zadanom prostorno-vremenskom okviru. služi za opis grupne mobilnosti i mjerenje socio-ekonomske aktivnosti u nekoj regiji, a najčešće se koristi u prometnoj znanosti za analizu i strateško planiranje prometnog opterećenja i prometne infrastrukture.

Osim tradicionalnog pristupa prebrojavanja putovanja i putnika na raskrižjima, anketiranjem ili pomoću različitih strateški postavljenih osjetila, za procjenu -e koristi se i statistička analiza podataka iz suvremenih informacijskih i komunikacijskih sustava (zapisi o aktivnostima u javnoj pokretnoj mreži, združena očitavanja prijamnika za satelitsku navigaciju i sl.). Takav suvremeni pristup procjeni omogućuje poboljšanje kvalitete procjene preslikavanjem -e na kontekst. Ako nosi informaciju o svrsi kretanja, specificira socio-ekonomske aktivnosti koje su pokretači putovanja, odnosno motivatori individualne i grupne mobilnosti.

Nedostaci kao što su mali uzorci stanovništva u anketama, neodređenost polazišta i odredišta putovanja, korištenje podataka koji nisu ažurni, čine tradicionalne metode za procjenu -e sve manje kompetentnima u rapidno promjenjivim urbanim sredinama.

Pojavom i razvojem velikog broja usluga zasnovanih na lokaciji, prikupljanje podataka o položaju u svrhu pružanja usluge omogućuje analizu kretanja stanovništva. Suvremeni izvori podataka o položaju i načini procjene -e koji uključuju kontekstu-

Poglavlje 1. Uvod

alizaciju su nov doprinos u području analize kretanja stanovništva.

Postojeće metode vrednovanja -e definiraju vrijednost nove matrice razmatranjem njene sličnosti s postojećom matricom za isto područje. Postoji potreba da se definira i kvantizira kvaliteta -e kroz objektivne parametre.

Rad započinje pregledom postupaka procjene -e i ukazuje na velike razlike u pristupima. Postupci su podijeljeni na tradicionalne i suvremene. Slijedi pregled postojećih postupaka vrednovanja -e baziranih na usporedbi -a i opis problematike vrednovanja -e suvremenih pristupa. Kroz pregled istraživanja opisana je uloga kontekstualizacije kao doprinosa u procjeni -e te je definiran alternativni pristup vrednovanju -e koji uvažava kontekstualizaciju. Predstavljena je metodologija usporedbe -a koju čine odnosni parametri kvalitete, pretpostavke usporedbe i postupak odlučivanja. Metodologija je demonstrirana na usporedbi dvije -e dobivene iz različitih izvora različitim postupcima procjene.

Poglavlje 2

Polazišno-odredišna matrica

Polazišno-odredišna matrica (*eng. Origin-Destination Matrix*, ODM ili *Trip Table*) je način mjerenja pokretljivosti stanovništva na nekom području u prostorno-vremenskom okviru. služi za opis grupne mobilnosti i mjerenje socio-ekonomske aktivnosti stanovništva nekog područja. Najčešće se koristi u prometnoj znanosti za analizu i strateško planiranje prometnog opterećenja i prometne infrastrukture [?]. može opisivati prometnu potražnju na jednom raskrižju, na području grada, regije, države i šireg područja. Povećanjem opisanog područja povećava se i zahtjevnost prikupljanja podataka za procjenu prometa i za izradu -e stoga postoji više postupaka prikupljanja podataka i modela za procjenu prometa.

2.1 Definicija polazišno-odredišne matrice

Definicija 2.1.1. Polazišno-odredišna matrica je matrica čiji svaki element predstavlja broj putovanja između odgovarajućeg para zona unutar promatranog područja u promatranom vremenskom okviru τ , u svakom smjeru zasebno. Redovi u -i $i = 1, 2, 3 \dots n$ predstavljaju polazišta (izvore) prometnih tokova, a stupci $j = 1, 2, 3 \dots n$ njihova odredišta (ušća). Element f_{ij} predstavlja broj putovanja iz pripadajuće polazišne zone i u pripadajuću ciljnu zonu j . Broj putovanja naziva se i širina prometnog toka (*eng. flow*). (slika ??)

Poglavlje 2. Polazišno-odredišna matrica

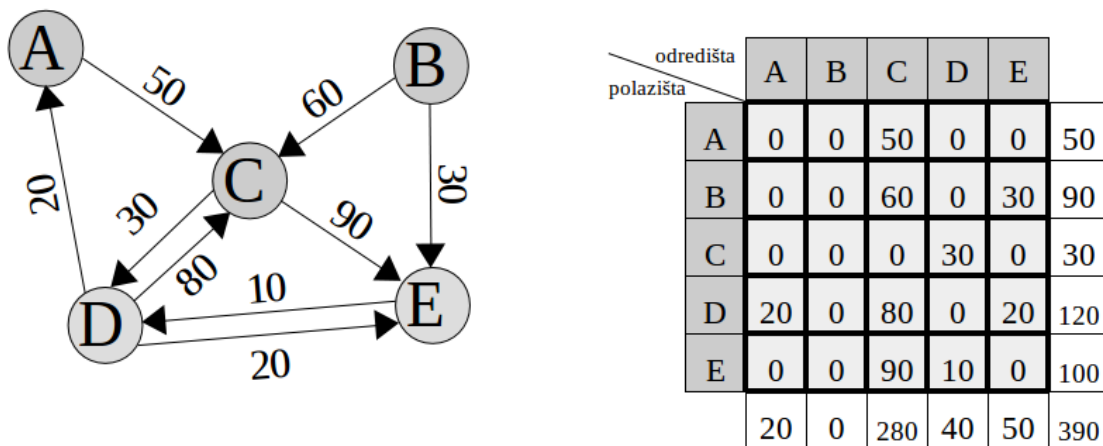
se može smatrati matričnim zapis težinskog, usmjerenog grafa, gdje su težinski faktori usmjerenih veza širine prometnih tokova između čvorova mreže [?]. Čvorovi predstavljaju zone na koje je promatrano područje podijeljeno (slika ??). Zone u -i su poredane od zapada prema istoku, te od sjevera prema jugu.

može biti procijenjena za sadašnje ili planirano prometno opterećenje [?]. Može biti prikazana ukupno ili po pojedinim načinima prijevoza (modalitetima prometnog sustava - osobna vozila, vozila javnog prijevoza, teretna vozila, pješaci itd.) i/ili različitim svrhama putovanja (putovanja na posao, putovanja kući...) [?]. Tranzitna (eng. *Transient Origin-Destination Matrix*, t-ODM) je vrsta -e u kojoj su zabilježene poznate tranzicije između zona. Svaka tranzicija predstavlja onaj dio putovanja koji je poznat, a početak i završetak cjelokupnog putovanja kojem pripada su nepoznati [?].

Poglavlje 2. Polazišno-odredišna matrica

Polazišta (Origins)	Odredišta (Destinations)							$\sum_{j=1}^n f_j = O_j$
	1	2	3	...	j	...	n	
1	f_{11}	f_{12}	f_{13}	...	f_{1j}	...	f_{1n}	O_1
2	f_{21}	f_{22}	f_{23}	...	f_{2j}	...	f_{2n}	O_2
3	f_{31}	f_{32}	f_{33}	...	f_{3j}	...	f_{3n}	O_3
...
i	f_{i1}	f_{i2}	f_{i3}	...	f_{ij}	...	f_{in}	O_i
...
n	f_{n1}	f_{n2}	f_{n3}	...	f_{nj}	...	f_{nn}	O_n
$\sum_{i=1}^n f_i = D_i$	D_1	D_2	D_3	...	D_j	...	D_n	$\sum_{i=1}^n \sum_{j=1}^n f_{ij} = F_{ij}$

Slika 2.1 Polazišno-odredišna matrica. Prema: [?]



Slika 2.2 Usmjereni graf kretanja i pripadajuća polazišno-odredišna matrica. (Zone su označene slovima abecede). Prema: [?]

2.2 Tradicionalni pristupi procjeni polazišno-odredišne matrice

Tradicionalni pristupi procjeni kao izvor podataka koriste podatke prikupljene anketom stanovništva i/ili prebrojavanjem vozila na strateškim položajima u prometnoj infrastrukturi.

2.2.1 Ankete

Tradicionalni pristupi procjeni -e uključuju metode provođenja anketiranja. Vrste anketa su: ankete u kućanstvima, presretanje vozila, anketiranje prijevoznih kompanija i tvrtki, tranzit na granici, turisti u hotelima i parking [?]. Ankete su zahtjevne iz perspektive utrošenog vremena, uloženog truda i radne snage stoga se ne provode često. Anketiranje putovanja u kućanstvima provodi se na područjima većine razvijenih svjetskih metropola tek **svakih 5 do 10 godina** [?]. Najveći postotak investicija u planiranju prometa odlazi na ankete, u prosjeku oko 7.4 milijuna dolara godišnje [?]. Ankete stanovništva bilježe broja stanovnika, broj vozila, broj zaposlenih, broj putovanja određenim načinom prijevoza (automobil, javni prijevoz...) po kućanstvu [?]. Ostale ankete pokrivaju malen dio populacije te su POM-e dobivene iz anketa opterećene neizbježnom pristranošću [?]. Studija na području Belgije pokazala je da se zadovoljavajuća razina kvalitete -e procijenjene iz anketa u kućanstvima postiže na uzorku veličine barem 50% populacije [?].

2.2.2 Prebrojavanje vozila

Metode prebrojavanja vozila (engl. *Traffic Counts*) na ključnim čvorištima prometne infrastrukture manje su zahtjevne za izvođenje u odnosu na ankete i smanjuju vrijeme i troškove prikupljanja podataka. Postupci uključuju: ručno prebrojavanje, analizu nadzornih snimaka prometa, osjetila prometnog toka (radarska, magnetna, video-senzori, zvučna), automatski sustav za prepoznavanje i očitovanje registarskih pločica (engl. , ANPR) [?]. **Navedenim postupcima nije moguće pouzdano odrediti stvarna polazišta i odredišta putovanja** [?], već se ona određuju

modeliranjem uz uvažavanje pretpostavki (sekcija ??). Točnost procijenjenih matrica ovisi o: korištenom modelu procjene, grešci ulaznih podataka i infrastrukturi detektora [?] [?]. Podaci dobiveni opremom koja radi na temelju vizualne percepcije nisu uvijek iskoristivi ako su vremenski uvjeti loši, a nadomjestak kod lošeg vremena je ručno prebrojavanje [?].

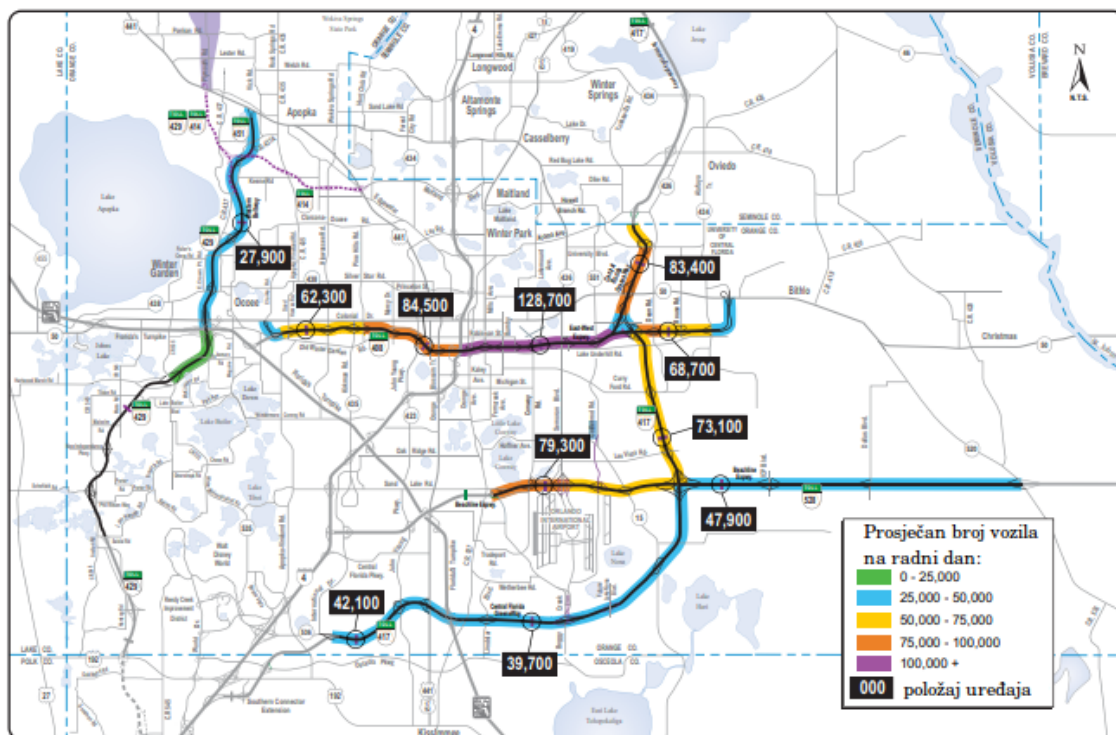
2.2.3 Raspored stanica za prebrojavanje

Stanica za prebrojavanje je položaja u infrastrukturi s kojeg se vrši prebrojavanje [?]. Raspored stanica treba biti planski, a broj dostupnih uređaja odnosno ljudi ograničavajući je faktor za veličinu područja koje je moguće nadzirati [?]. Točnost -e povećava se s brojem stanica [?]. Uređaji na različitim položajima u prometnoj mreži imaju različit stupanj utjecaja na procjenu -e [?]. Važno je odrediti optimalan broj stanica prebrojavanja i njihovih položaja u prometnoj mreži. Optimalni uvjeti će se dobiti poštujući četiri pretpostavke [?]:

1. Najmanje jedna stanica prebrojavanja po paru zona.
2. Stanice trebaju biti na prometnim vezama (cestama) gdje protječe najveći mogući prometni tok između te dvije zone.
3. Odabrana prometna veza treba presijecati što više tokova.
4. Rezultati prebrojavanja prometa na svim vezama trebaju biti linearno nezavisni.

Prva pretpostavka neće biti ispunjena ako nisu osigurana sredstva [?]. Prošireni modeli za raspodjelu stanica u današnjim programskim podrškama imaju mogućnost ograničavanja investicija u infrastrukturu za promatranje. Model prometne potražnje može imati 2,000 , što znači da pripadajuća -a ima 4 milijuna ćelija, dok je u stvarnosti na prometnoj mreži postavljeno svega 1,000 stanica za prebrojavanje [?]. Na slici ?? prikazana je karta grada Orlando s važnim prometnicama (sivom) i s prosječnim prometnim tokovima zabilježenim uređajima za prebrojavanje vozila. [?]

Poglavlje 2. Polazišno-odredišna matrica



Slika 2.3 Prikaz prosječnog broja vozila na vezama (prometnicama) sa stanicama za prebrojavanje radnim danom 2005. Orlando, Florida. [?]

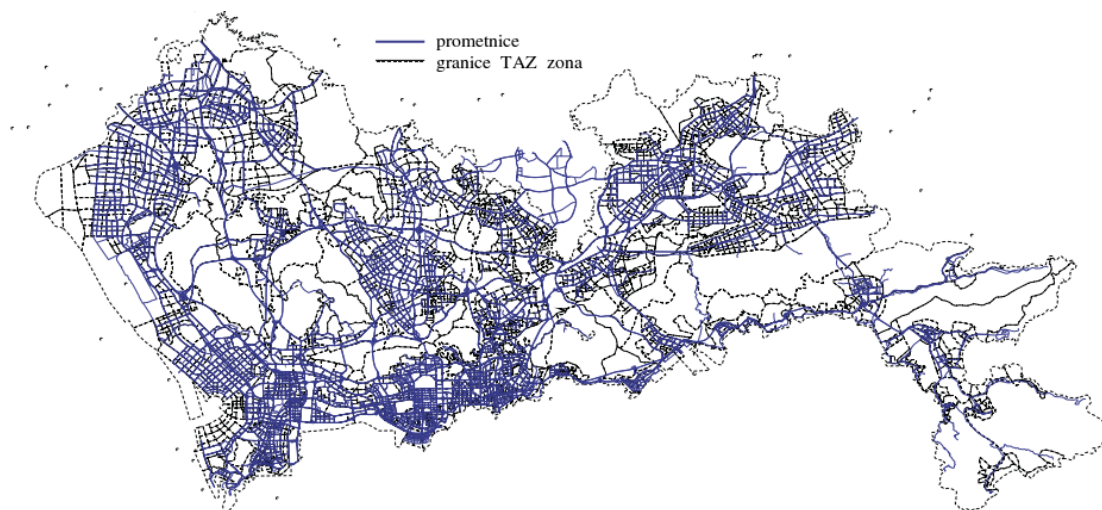
2.2.4 Prometne zone

Prometna zona je službena jedinica prostorne, geografske podjele (zone) kod analize prometa [?]. Strani naziv za prometnu zonu je *Traffic Analysis Zone*. Površina tih jedinica varira i nije uniformna. Podjela na zone se može temeljiti na podjednakoј populaciji, podjednakoј broju zaposlenih ili podjednakoј broju izlaznih putovanja (jednaka širina izlaznih tokova)[?]. Primjerice, svaka zona pokriva prostor na kojem obitava otprilike tri tisuće ljudi [?].

U hijerarhijskoј podjeli kod ankete stanovništva postoje manje i veće jedinice podjele, a nema podjele koja bi precizno odgovarala [?]. Anketni blokovi određeni su kao relativno homogene jedinice kad je riječ o karakteristikama populacije, ekonomskom statusu i uvjetima života u vrijeme kada su uspostavljeni [?]. Do 2000. godine zastupljen je bio samo jedan model predviđanja prometa, a

veličina bila je približno jednaka anketnom bloku od 600 do 3,000 ljudi [?]. Pojavom kompleksnijih modela predviđanja prometa počinju se koristiti sve manje prostorne jedinice jer se povećava i razina detalja koju noviji modeli mogu ostvariti [?]. Jedinica veća od uvedena 2010. naziva se Traffic Analysis District (TAD) i obuhvaća oko 20,000 stanovnika, a koristi se za analize na razini države. [?]

Kineski grad Shenzhen ima preko 10 milijuna stanovnika i proteže se na 1,991 km^2 [?]. Istraživanje objavljeno 2017. koristi podjelu na 3,199 (Vidi sliku ??). Zaključak iz danih informacija je: prosjek površine u gradu Shenzhenu 0.62 km^2 i u prosjeku jedna zona obuhvaća više od 3,125 stanovnika.



Slika 2.4 Primjer mreže za analizu prometa 221,14 cesta - plave linije i 3,199 zona - granice označene iscrtanom crnom linijom [?]

2.2.5 Model prometa

Modeliranje prometa zahtjeva veliku količinu podataka - informacija o prometnoj mreži i prometnoj potražnji. Točnost modelirane (planirane) prometne situacije ovisi o kvaliteti dostupnih informacija te kako su podaci kombinirani, koji težinski faktori su primijenjeni za različite izvore [?]. Prometna potražnja ključna je komponenta te stoga gotovo svaki prometni model zahtjeva -u koja specificira prometnu potražnju

između zona u prometnoj mreži [?]. Točna osnova je za odluke kod strateških planiranja prometne infrastrukture i u [?] koji su zauzvrat ključni za izbor ruta u različitim sustavima navigacijskog navođenja.

Primjeri pretpostavki u modeliranju prometne potražnje su: ukupna dnevna potražnja podjednaka u oba smjera kod svakog para zona na području grada [?], korisnici imaju saznanja o situaciji na cesti [?], mobilnost manja zimi nego ljeti [?] i broj putovanja po kućanstvu proporcionalan dohotku [?]. Klasičan bazni model za procjenu prometne potražnje je gravitacijski model (eng. *Gravity model*) koji kod proračuna "privlačnosti" 2 zone uzima u obzir **gustoću populacije, udaljenost i troškove putovanja**. *Broj putovanja između dvije zone srazmjeran je veličini aktivnosti odnosno nastajanju i privlačenju putovanja u tim zonama, a obrnuto proporcionalan prostornoj udaljenosti između zona ili nekoj drugoj vrsti otpora (vrijeme putovanja, pješaćenja, čekanja, parkiranje, vozarina ili kombinacija svih navedenih otpora)* [?]. Izvedene verzije gravitacijskog modela integriraju i neke druge (socio-ekonomske) faktore [?].

Procjena iz prebrojanih vozila je matematički pododređena (više je nepoznanica nego poznatih informacija) [?]. To znači da bi velik broj matrica odgovarao tako prebrojanim tokovima vozilima. **Prebrojavanje i ankete se kombiniraju u procjeni POM-e**. Temeljna ideja kod modeliranja je [?]:

1. Pronaći postojeću POM-u za isto područje i raspodjelom njenih tokova na prometnu infrastrukturu simulirati rezultate prebrojavanja na odabranim točkama u prometnoj mreži [?].

Ako postoji više POM-a za isto područje potrebno odabrati onu koja najviše odgovaraju za konkretni slučaj. Većina algoritama koristi početnu (engl. *seed*) POM-u koja je *najbolja procjena* dosadašnje potražnje, a koja procjenjuje prometnu potražnju u prošlosti (npr. iz ankete po kućanstvima), nedavnu prometnu potražnju neprecizno ili je potražnja razvijena iz *principa tipičnog ponašanja vozača* [?].

2. Početnu -u modificirati iterativnim procesom evaluacije omjera tokova dodijeljenih prometnim vezama iz -e i prebrojenih tokova na vezama gdje se promet prebrojavao [?].

U tom iterativnom procesu prometni tokovi iz modificirane -e ponovno se dodjeljuju prometnoj infrastrukturi (sve dok se ne postigne zadani broj iteracija ili zadovoljavajući stupanj sličnosti početne i nove -e - detaljnije u sekciji ??).

2.2.6 Vremenski nezavisno i vremenski zavisno modeliranje

Modeliranje prometa dijeli se na dvije grane: vremenski nezavisnu (statičku, analitičku) i vremenski zavisnu (simulacijsku) [?]. **Cilj statičkog modeliranja je analiza prometa za strateška, investicijska planiranja izgradnje nove infrastrukture. Modeli procjenjuju prosječnu dnevnu prometnu potražnju (period od 16 ili 24 sata) [?] . Radi se o vremenski nezavisnoj analizi.**

Vremenski zavisno modeliranje koristi se kod upravljanja prometnom signalizacijom i definiranja rasporeda vožnje javnog prijevoza. Period u kojem se procjenjuje prometna potražnja manji je od jednog dana, primjerice period τ od 15 minuta [?][?]. Riječ je o vremenski zavisnoj analizi prometa.

Neki *Peak-hour* vremenski modeli koji opisuju potražnju za vrijeme dnevnog vrhunca prometne potražnje, procjenjuje potražnju za svaku minutu ($\tau = 1 \text{ min}$) [?]. Modeliranje u stvarnom vremenu svrstava se u vremenski zavisno modeliranje .

Većina razvijenih modela za vremenski zavisno modeliranje iz postupka prebrojavanja (engl. *link count*) primjenjuju se na malim, zatvorenim prometnim mrežama kao što su raskrižja, petlje i manje autoceste [?].

2.3 Suvremeni pristupi procjeni polazišno-odredišne matrice

Podatci koje generiraju pametni telefoni i pametne kartice omogućuju podatkovnim analitičarima razumjeti ponašanje pojedinaca u različitim domenama, uključujući i njihov obrazac kretanja. Informacijske i telekomunikacijske usluge generiraju masivne skupove podataka vezanih za gibanje u prostoru (mobilnost). Danas ti podaci postaju suvremeni izvori podataka o dnevnim mobilnosti stanovništva. Podaci iz tih suvremenih izvora nisu ciljano prikupljeni za analizu mobilnosti stanovništva već

za pružanje usluge zasnovane na lokaciju te se koristi naziv pasivno prikupljani polazišno-odredišni podaci (engl. *passively collected origin-destination data*) [?] [?]. **Suvremeni izvori podataka o položaju mogu ili ne moraju biti vezani uz prometnu infrastrukturu niti za određen način kretanja.**

Pružatelji usluga zasnovanih na lokaciji i operatori javne pokretne mreže moraju voditi računa o privatnosti pojedinaca kada je riječ o podacima koji se prikupljaju, uključujući podatke o položaju korisnika. Ukoliko ustupe podatke za razvoj , ti podaci moraju biti anonimizirani te ne smije postojati mogućnost povratnog inženjerstva kojim bi se povratio izvorni identitet. Važno je osigurati anonimnost i dozvolu pojedinca da se podaci koriste u jasno definirane svrhe. Idealno je da operator ustupi već anonimizirane podatke stranci koja iz njih procjenjuje -e.

2.3.1 Pokretna osjetila

Pokretno osjetilo je uređaj koji prikuplja opaža vrijednosti definiranih parametara gibanja vozila na koje je postavljeno te ih pohranjuje i prenosi radi naknadnog procesiranja [?]. Vozilo se kreće prometnom infrastrukturom, postajući tako dijelom prometnog toka. Osnovni podatci koje osjetilo prikuplja su položaj, brzina i smjer kretanja vozila [?].

Nužni zahtjev je poznavanje položaja vozila. Uređaj određuje položaj pomoću prijemnika globalnog navigacijskog satelitskog sustava () ili se vozilo identificira na određenim točkama u prometnoj mreži, na kojima može ujedno i prenijeti podatke prikupljene od posljednje točke [?]. Anglosaksonski naziv za ovu metodu prikupljanja podataka o stanju u prometu je . metoda zastupljena je u . **U prikupljanju podataka sudjeluju uređajima opremljena taksi vozila, službena vozila, vozila za distribuciju robe i vozila javnog prijevoza** [?]. ITS koji zahtijeva podatke u realnom vremenu mora imati uspostavljen komunikacijski sustav za učestali prijenos podataka o položaju. Prijenos podataka najčešće se odvija preko , Wi-fi i Bluetooth komunikacijskih sustava [?].

Ukoliko pokretno osjetilo nije specijalizirani uređaj već pametni telefon koristi se naziv (FPD) ili *Floating Cellular Data* [?]. se odnosi na prikupljanje podataka pomoću mobilnih (pametnih) telefona neovisno o načinu određivanja položaja -

korištenjem -a ili pomoću javne pokretne mreže [?]. Položaj pokretne stanice unutar javne pokretne mreže može se odrediti na dva načina. Prvi je triangulacija signala koje pokretna stanica hvata iz više baznih stanica. Drugi je izjednačavanje položaja pokretne stanice s položajem (područjem pokrivanja) bazne stanice s kojom je pokretna stanica povezana u trenutku očitavanja položaja [?] [?].

Osnovni postupak određivanja položaja pomoću prijamnika sadržani neke pogreške. Točnost u rangu od 10-20 m je moguća uz dostupnost signala od najmanje 4 satelita. U urbanom okruženju zgrade i visoke strukture često blokiraju vidljivost između satelita i prijamnika. Rezultat toga je degradacija točnosti određenog položaja. Napredno određivanje položaja podrazumijeva usporedbu i usklađivanje inicijalno određene putanje kretanja s kontekstom (npr. poznatom infrastrukturom) (engl. *map matching*)[?]. Njime se postiže točnost u rangu 0-15 m [?][?].

2.3.2 Polazišno-odredišna matrica iz zapisa o telekomunikacijskim aktivnostima u javnoj pokretnoj mreži

U postupku praćenja aktivnosti za potrebe pružanja i naplatu usluga te za održavanje operativnosti same mreže zapisuju se podaci o aktivnostima u javnoj pokretnoj mreži. Pokretni uređaj (npr. pametni telefon) je kontinuirano povezan s javnom pokretnom mrežom preko odgovarajuće bazne stanice, neovisno o telekomunikacijskoj aktivnosti uređaja. U postupku uspostave poziva, signalizacija uključuje i identifikacijsku oznaku bazne stanice s koje je poziv uspostavljen. Navedena oznaka bilježi se u zapisu o telekomunikacijskoj aktivnosti. Zapisi o telekomunikacijskoj aktivnosti i detaljima telekomunikacijskih transakcija vode se u odgovarajućim komponentama javne pokretne mreže pod imenima: . CDR obuhvaća transakcijske podatke o pozivima, razmjeni kratkih poruka (engl. *Short Messaging System*,) i prijenosu podataka pokretnim internetom. Transakcijski podaci uključuju između ostalog: vremenske trenutke početka i završetka telekomunikacijske transakcije (korištenja komunikacijske usluge), vrstu komunikacije, identifikacijske oznake pošiljatelja i primatelja te identifikacijske oznake baznih stanica na koje su u trenutku započinjanja i završetka transakcije bili spojeni pošiljatelj i primatelj (vidi tabele ?? i ??)

Poglavlje 2. Polazišno-odredišna matrica

Tablica 2.1 Primjer zapisa o telekomunikacijskim aktivnostima (pozivi) [?]

ID korisnika	Vrijeme događaja	Bazna stanica	pozivatelj	primatelj	Dolazni/Odlazni	Trajanje (s)
38DA6	2015-05-01 18:26:50	1921	38DA6	163B7	Odlazni	52
78EC3	2015-05-01 14:16:09	2189	53880	78EC3	Dolazni	600

Tablica 2.2 Primjer zapisa o telekomunikacijskim aktivnostima [?]. Uklonjen je datum, a identifikacijska oznaka bazne stanice (pozivatelja) zamijenjena je njenim položajem. Podatci o baznim stanicama dostupni su u bazama poput OpenCellID¹

ID SIM kartice	Vrijeme	Geografska dužina	Geografska širina
0055555556	13:10:09	114.0397917	22.5740278
0055555556	16:08:30	113.8056944	22.7613889
0055555556	20:50:17	113.8465972	22.7725
0055555557	23:39:23	114.043854	22.530757
0055555558	22:13:27	114.043854	22.530757

Zapisuju se i podatci o prekapčanju pokretnog uređaja na drugu baznu stanicu (engl. *handover data*) tijekom trajanja telekomunikacijske aktivnosti radi postizanja i održavanja potrebne kvalitete telekomunikacijske usluge, podatci koju skupinu baznih stanica je pokretna stanica spojena, kada je pokretna stanica uključena/isključena iz mreže i opterećenje pojedinog tornja [?]. Svaka skupina baznih stanica ima od 150 do 500 baznih stanica [?]. Podatci o skupini na koju je pokretna stanica spojena (engl. *Location Area Updates*) prikupljaju se svakih 20 min do 2 sata neovisno o aktivnosti [?].

Novi postupci koriste upravo anonimizirane, često višemjesečne, kao izvor poda-

¹ zapisi ne sadrže uvijek položaje baznih stanica već samo njihov ID. Postoji nekoliko javno dostupnih baza kao što je OpenCellID [?] nastalih volonterskim doprinosom koje pohranjuju podatke o procjenama položaja baznih stanica, snazi signala i sl. Položaji se prikupljaju pomoću aplikacije instalirane na mobilnom uređaju, registriranjem položaja i identifikacijske oznake bazne stanice na koju je uređaj trenutno spojen.

taka za procjenu -a [?][?][?]. Metodologija korištenja odnosnih elemenata CDR-a za procjenu POM-e predložena je još 1994. godine [?].

Definicija 2.3.1. Zaustavljanje je skup minimalno n događaja (telekomunikacijskih aktivnosti) u nizu istog korisnika koji su zabilježeni na istom položaju unutar definiranog minimalnog vremenskog perioda (engl. *Stop*) .

Zahtjevom da se radi o više ($n > 1$), a ne jednom ($n = 1$) zabilježenom događaju pokušava se osigurati da se ne radi o tranzitnom već o stvarnom zaustavljanju i suzbija se bilježenje lažnog kretanje kod privremenog spajanja na susjednu baznu stanicu [?]. Ako je $n > 1$ zaustavljanje ima 2 vremenske oznake, trenutak kada je zabilježen početak i trenutak kada je zabilježen kraj zaustavljanja.

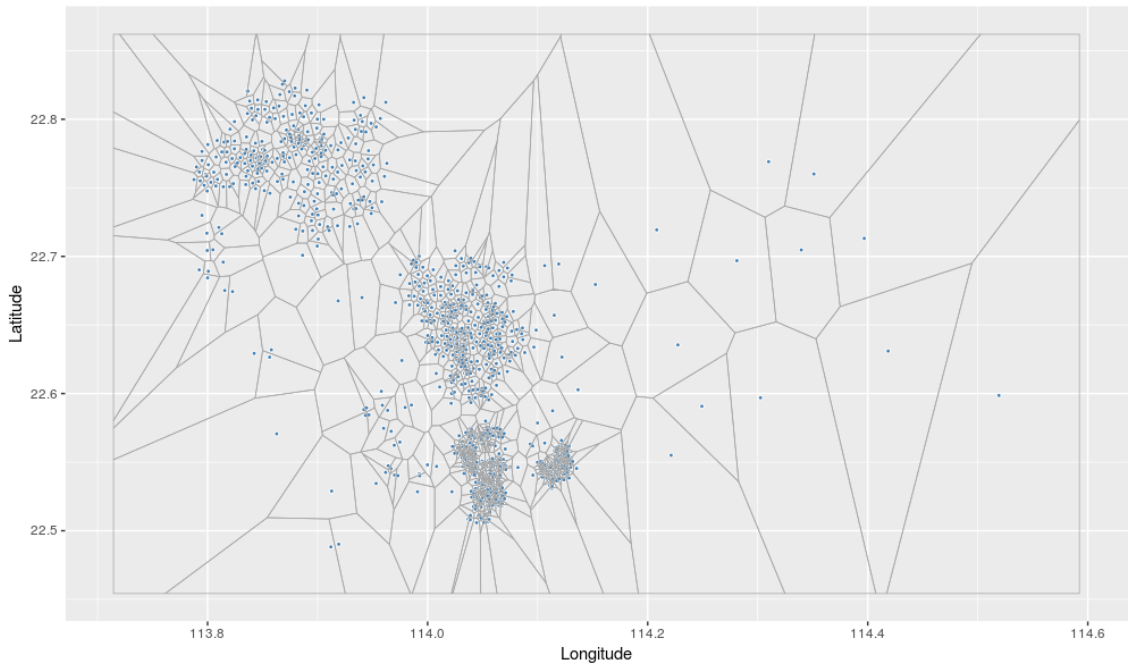
Definicija 2.3.2. Putovanje je skup dva zaustavljanja koja zadovoljavaju unaprijed definirana minimalna i maksimalna ograničenja međusobne vremenske i prostorne udaljenosti. Kraj polazišnog zaustavljanja postaje početak putovanja, a početak odredišnog zaustavljanja kraj putovanja. Niz kontinuiranog zaustavljanja na različitim položajima otkriva putanju kretanja.

2.3.3 Prostorna podjela - Voronoi teselacija

Podjelom prema Voronoiovoj teselaciji postiže se *najfinija* prostorna rezolucija -e iz .

Definicija 2.3.3. Voronoi teselacija definirana je skupom točaka P u ravnini, gdje za svaku točku p_i Voronoi ćeliju K_i čine sve točke ravnine bliže toj točki negoli ikojoj drugoj točki iz skupa [?]. Svaka takva ćelija dobiva se presjecištem pola prostora između susjednih točaka skupa, pa stoga čini konveksni poligon. Dijelovi Voronoiova dijagrama su sve točke u ravnini koje su ekvidistantne prema dva najbliža položaja (granice poligona). Voronoiovi vrhovi su točke ekvidistantne tri (ili više) položaja. [?]

Točke skupa su položaji baznih stanice, a ćelije oko njih aproksimacija površine koju ta bazna stanica pokriva (vidi sliku ??). Površine prekrivanja u stvarnosti se preklapaju jer područje prekrivanja nije sasvim fiksno i ovisi o trenutnoj aktivnosti svake bazne stanice [?].



Slika 2.5 Voronoi dijagram 1090 ćelija baznih stanica - kineski grad Shenzhen [?]

Gustoća Voronoi ćelija dobro je korelirana s gustoćom stanovništva te stoga bolju prostornu rezoluciju ima na područjima veće gustoće populacije [?]. Poistovjeti li se položaj korisnika s položajem tornja bazne stanice, mjerna nesigurnost određivanja položaja iznosi od približno stotinu metara u gusto naseljenom, do nekoliko kilometara u ruralnom području [?]. Problem predstavlja preklapanje područja pokrivanja u stvarnosti [?]. Zbog uravnoteženja opterećenja u mreži (eng. *load balancing*) uređaj mijenja baznu stanicu na koju je spojen iako sve vrijeme zadržava isti položaj. Učinak koji uravnoteženje opterećenja ima na definiranje putovanja može se ublažiti određivanjem minimalne udaljenost putovanja, zanemarivanjem putovanja u susjedne ćelije [?] ili agregacijom područja pokrivanja baznih stanica čija je udaljenost manja od 75 m (s dodjelom događaja najaktivnijem tornju) [?].

Analiza snage signala i trenutnih opterećenja okolnih baznih stanica mogu dati jasniji uvid u opterećenje mreže u vremenu početka (i trajanja) telekomunikacijske aktivnosti [?].

2.3.4 Diskretizacija vremena

Suvremeni izvori podataka o kretanju bilježe vremenski žig uz podatak o položaju što ih čini pogodnima za vremenski zavisnu analizu. Diskretizacija vremena je praktički neizbježna zbog ograničene granularnosti [?]. U literaturi je korištena različita vremenska rezolucija (razlučivost) - podjela na vremenske periode τ od 15 min, 30 min [?], 1 sat, 2 sata, 3 sata [?][?]. [?, ?] Putovanja ili zaustavljanja dodijeljena su periodima u kojima se odvijaju. Još jedna od zastupljenih podjela je na specifične periode - *radni sati*, *noćni sati*, *komutacijski sati* [?][?][?] u kojoj se promatra gdje pojedinci pretežito borave u tipičnom radnom vremenu te tokom noći kroz cijelo razdoblje prikupljanja, \mathcal{T} .

2.3.5 Rekonstrukcija putanja

Kretanje pojedinca tokom dana najčešće je rascjepkano, odnosno u nizu postoje "rupe" kada je informacija o kretanju nepoznata zbog nepostojanja aktivnosti u tom vremenu.

Definicija 2.3.4. Potpunost putanje definirana je izrazom [?]:

$$potpunost = \frac{n_L}{\tau \times \mathcal{T}} \quad (2.1)$$

gdje je:

n_L broj zabilježenih položaja (zaustavljanja)

τ vremenski period

\mathcal{T} duljina ukupnog vremenskog razdoblja za koje su dostupni zapisi.

Idealna potpunost je postignuta ako postoji barem jedan zabilježeni položaj (zaustavljanje) po periodu [?].

Veća potpunost putanja postiže se ako se osim poziva kao komunikacijskih aktivnosti u obzir uzmu poruke i prijenos mobilnih podataka [?]. **Odbacivanje korisnika s niskim stupnjem potpunosti** zastupljeno je u istraživanjima razvoja -a iz što dovodi do iskorištavanja tek vrlo malog dijela korisnika - tek 1.67% i 0.45% u nekim istraživanjima [?].

Poglavlje 2. Polazišno-odredišna matrica

Rekonstrukcija putanja na osnovu aktivnosti u cijelom vremenskom razdoblju promatranja \mathcal{T} (npr. 7, 15, 30, 90 dana) nameće se kao novi postupak koji povećava broj i točnost rekonstruiranih putanja i smanjuje mogućnost povratnog inženjeringa identiteta . Rekonstrukcijom putanja dolazi do povećanja broja istih putanja (više pojedinaca) koje se odvijaju isto vrijeme, čime se štite pojedinci koji prije postupka jedini imaju veliku potpunost putanja [?].

Poglavlje 3

Tradicionalni pristupi vrednovanju polazišno-odredišnih matrica

Postupak vrednovanja je objektivna procjena kvalitete, vrijednosti objekta koji se vrednuje. Pregledom literature o vrednovanju -e utvrđeno je da se jednoznačno koriste pojmovi točnost, pouzdanost i kvaliteta -e. Vrednovanje -e vrši se postupcima koji koriste mjere razine sličnosti odnosno razlike (greške) s referentnom matricom -a temelj je razlika na razni svakog elementa -e.

Tradicionalni postupak vrednovanja -e svodi se na određivanje njene sličnosti sa istinitom (engl. *ground truth*) POM-om [?]. Istinita -a je POM-a uzeta kao najtočniji opis stvarnog stanja. Istinita -a dobivena je tradicionalnim postupcima (anketiranje i/ili prebrojavanje vozila) [?]. Statističke mjere kvantiziraju razliku procijenjenih i pravih vrijednosti (vidi dodatak ?? - Metrike za vrednovanje polazišno-odredišne matrice). Iz statističkih mjera sličnosti izdvojen je korelacijski koeficijent određenja R^2 koji je opisan u narednoj sekciji.

Upravo ta referentna istinita POM je ujedno i početna (iz sekcije ??) u postupku procjene nove -e iz rezultata prebrojavanja vozila [?]. će biti vrednovana usporedbom s -om koja se koristi u postupku procjene te iste -e. Postupak vrednovanja nove -e stoga je ovisan o kvaliteti početne -e, koja je procijenjena u odnosu njenu početnu matricu i tako rekurzivno.

3.1 Linearna regresija - Korelacijski koeficijent određenja R^2

Tehnika vrednovanja linearnom regresijom procjenjuje vrijednosti ćelija jedne -e (zavisna varijabla - y) na temelju vrijednosti druge - referentne istinite -e (nezavisna varijabla - x). Vrijednosti u -i promatraju se kao skup vrijednosti $i = 1, 2, 3, \dots, n^2$ ne uzimajući u obzir njihovu matricnu strukturu. Pretpostavka kod linearne zavisnosti je da postoje koeficijenti a i b takvi da se za svaku vrijednost nezavisne varijable x_i zavisna varijabla y_i može napisati kao [?]:

$$y_i = bx_i + a + \epsilon_i \quad (3.1)$$

gdje je :

pogreška ϵ slučajna varijabla s normalnom raspodjelom i očekivanom vrijednošću 0

b koeficijent smjera pravca linearne regresije

a slobodni član (predstavlja križište linije regresije s y osi).

Linija linearne regresije prikazana je na slici ??.

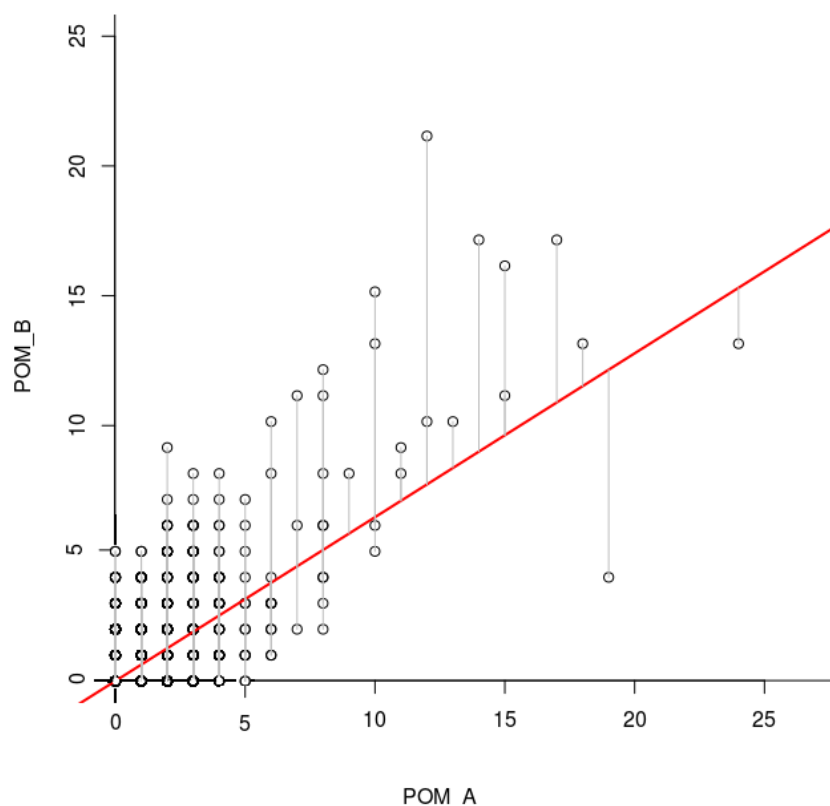
Korelacijski koeficijent određenja ili obuhvaćanja varijance R^2 je mjera koja pokazuje koliko je slaganje između vrijednosti izračunatih modelom procjene i izmjerenih, empirijskih vrijednosti . Općenito, to je mjera dobrote pristajanja (engl. *goodness of fit*) koja govori koliko uspješno model opisuje zavisnost (korelaciju) između zavisne i nezavisne varijable [?]. R^2 predstavlja omjer modelom opisanih varijacija u odnosu na ukupne varijacije podataka [?]:

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum (y_i - \bar{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (3.2)$$

Raspon mjere R^2 je od 0 do 1, 0 ukazuje na nepostojanje objašnjenja varijance između izračunatih i izmjerenih podataka, a 1 ukazuje da je varijanca u potpunosti objašnjena modelom.

Ako su matrice potpuno iste, vrijednost $R^2 = 1$, $a = 0$ i $b = 1$. Matrice su sličnije što su vrijednosti bliže navedenima [?]. Kod usporedbe dvije POM-e u odnosu na referentnu istinitu POM-u, ona koja je sličnija referentnoj bolje opisuje kretanje.

Zadovoljavajuća vrijednost R^2 u usporedbi s referentnom istinitom POM-om navodi se 0.99 [?]. Ta vrijednost ne smatra se praktičnom za usporedbu POM-a dobivenih različitim pristupima procijeni što dolazi do izražaja i kada se radi o -i iz suvremenih izvora [?].



Slika 3.1 Primjer linearne regresije između vrijednosti dvije POM-e. Vrijednosti POM-e A su na x osi, a vrijednosti POM-e B su na y osi. Pravac linearne regresije označen je crvenom linijom, reziduali sivim linijama.
 $a = 0.0071273$, $b = 0.631128$, $R^2 = 0.4026$

3.2 Vrednovanje suvremene polazišno-odredišne matrice

Unaprjeđenje metodologije procjene -e koja će pružiti povećanu kvalitetu i transparentnost uz ispunjavanje suvremenih zahtjeva na tehnologije (cijena, zaštita privatnosti i podataka itd.) privlači pažnju i van područja prometne znanosti. Suvremeni pristupi koriste alternativne izvore podataka o kretanju za promatranje obrazaca putovanja i pristupe procjeni -e, no prema [?] nailaze na problem da prometna struka zahtijeva da se njihova vrijednost dokazuje usporedbom s tradicionalnim pristupima, dobro poznatim u praksi. Za tradicionalnu usporedbu suvremene -e čija prostorna podjela nije s -om čija je prostorna podjela na traži da se prostorna podjela prilagodi nekoj novoj ili prostornoj podjeli jedne od matrica te da se nakon toga matrice usporede *elementwise*. Prilagodba se u praksi čini agregacijom i/ili interpolacijom tokova. (Detaljnije u Dodatku ?? - Prilagodba prostorne podjele.)

3.3 Grafički oblik i sličnost strukture polazišno-odredišne matrice

Vrijednosti u matrici mogu se normalizirati na opseg vrijednosti koji je moguće prikazati u obliku slike tako da svaka vrijednost predstavlja vrijednost piksela. Različite palete korištene su za prikaz raspona, od raspona 0-100 koji predlažu [?], prikazom nijansama jedne boje 0-255 (engl. *grayscale*) [?], paletom nijansa 2 boje, ili kojom drugom proizvoljnom paletom s više boja .

Nekolicina autora ističe važnost strukturalne sličnosti s referentnom matricom kao važnu mjeru kvalitete -e jer visoka razina strukturalne sličnosti može biti prisutna i kod -a s manjom razinom sličnosti prema statističkim mjerama [?][?][?]. Također, strukturalna sličnost je (vizualno) vidljiva u grafičkom obliku -e. **Strukturalna sličnost dobro odgovara ljudskoj vizualnoj percepciji sličnosti slika [?].**

3.3.1 Višerazinski indeks strukturalne sličnosti

Višerazinski indeks strukturalne sličnosti (engl. , MSSIM) dolazi iz područja računalne obrade slike i koristi se kao mjera usporedbe digitalnih slika (*eng. measure of comparison*). Ideja o korištenju MSSIM u analizi urbanog kretanja stanovništva za mjerenje sličnosti -a prvi puta se spominje i demonstrira na simuliranim -ma dobivenim iz referentne -e dodavanjem šuma [?].

Informacija o strukturi slike definira se kao atributi slike koji predstavljaju strukturu objekata na sceni, i neovisni su o prosječnom osvjetljenju i kontrastu. Jer osvjetljenje i kontrast mogu znatno varirati na sceni, moraju se u obzir uzeti samo njihove lokalne vrijednosti [?]

Indeks strukturalne sličnosti (engl. , SSIM) se bazira na degradaciji strukturalnih informacija na jednoj slici u usporedbi s drugom (referentnom) slikom. se računa za svaki kvadratni blok veličine N elemenata na način da se jezgra (da bi obuhvatila novi blok) pomiče ćeliju po ćeliju dok ne prođe preko cijele slike. je srednja vrijednost svih .

3.3.2 Osnovni višerazinski indeks strukturalne sličnosti

Neka su X i Y matrice koje uspoređujemo a $x = \{x_n | x = 1, 2, \dots, N\}$ i $y = \{y_n | y = 1, 2, \dots, N\}$ parovi vrijednosti kvadratnih prozora veličine jezgre na istim pozicijama u X i Y ; $SSIM$ je određen prosječnim vrijednostima μ_x i μ_y sa standardnim devijacijama σ_x i σ_y i kovarijancom σ_{xy} [?]

$$l(x, y) = (2\mu_x\mu_y + C1)/(\mu_x^2 + \mu_y^2 + C1) \quad (3.3)$$

$$c(x, y) = (2\sigma_x\sigma_y + C2)/(\sigma_x^2 + \sigma_y^2 + C2) \quad (3.4)$$

$$s(x, y) = (\sigma_{xy} + C3)/(\sigma_x\sigma_y + C3) \quad (3.5)$$

Izraz $l(x, y)$ opisuje razliku u osvjetljenju, $c(x, y)$ razliku u kontrastu, a $s(x, y)$ razliku u strukturi između x i y . $C1$, $C2$ i $C3$ su konstante uvedene da se izbjegne "nestabilnost" kada su nazivnici bliski 0. Opća forma SSIM definira se kao [?]:

$$SSIM(x, y) = [l(x, y)^\alpha][c(x, y)^\beta][s(x, y)^\gamma] \quad (3.6)$$

gdje su α , β i γ parametri relativne važnosti svake komponente. Za SSIM vrijedi sljedeće:

$$SSIM(x, y) \leq 1 \quad (3.7a)$$

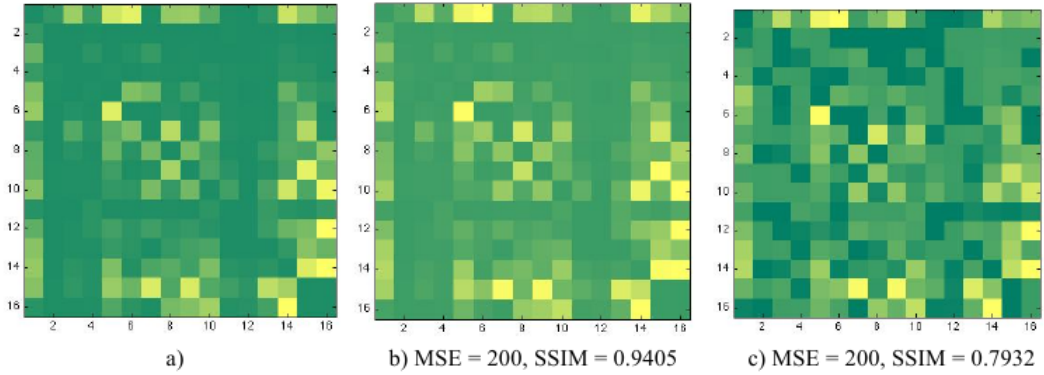
$$SSIM(x, y) = SSIM(y, x) \quad (3.7b)$$

$$SSIM(x, y) = 1 \iff x = y \quad (3.7c)$$

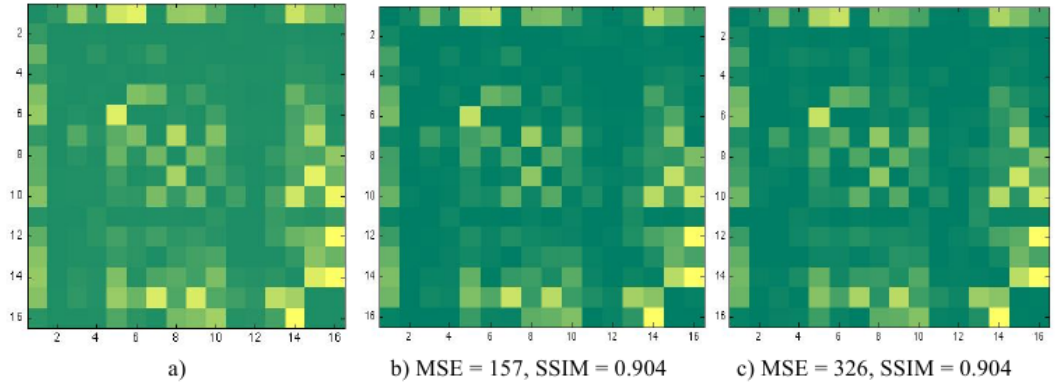
$$MSSIM(X, Y) = \frac{1}{M} \sum_{m=1}^M SSIM(x_m, y_m) \quad (3.8)$$

3.3.3 Poboljšani višerazinski indeks strukturalne sličnosti

Nekoliko godina nakon prvog spominjanja MSSIM kao metrike usporedbe -a Vuren [?] se dotiče 3 problema postavljajući pitanja: Koliko treba biti veliki blok? Kako usporediti "guste" i "rijetke" -e? Koja je prihvatljiva vrijednost MSSIM? Autori definiraju poboljšani model koji nazivaju 4D-MSSIM u izračun kojeg dodaju stvarne euklidske udaljenosti prostornih zona. Tehnike bazirane na strukturalnoj sličnosti i njihov doprinos još su uvijek otvorene za raspravu [?][?][?].



- (a) Vrijednost metrike točnosti MSE jednak je za obje -e b) i c) u odnosu na referentnu istinitu -u a). Indeks strukturalne sličnosti se znatno razlikuje, gdje b) ima veću strukturalnu sličnost s a) nego što ima c). Podsjetimo se da za $SSIM(a, a)$ vrijedi (??). Lako je na matrici c) *golim okom* uočiti narušenost strukturalne sličnosti.



- (b) Vrijednost metrike strukturalne sličnosti iznosi 0.904 za obje -e b) i c) u odnosu na referentnu -u a), dok se MSE razlikuje (veći MSE znači veću razliku). Objе -e dobivene su dodavanjem Gaussianovog šuma na referentnu matricu.

Slika 3.2 Narušenu strukturalnu sličnost lakše je uočiti nego narušen MSE [?].

Poglavlje 4

Na kontekstualizaciji zasnovan pristup vrednovanju polazišno-odredišnih matrica

4.1 Uvodna razmatranja

Širenje gradova i rast stanovništva rezultirali su rastućim brojem sve ozbiljnijih prometnih zastoja u gradovima diljem svijeta. Strategije upravljanja prometom i gradom traže uvažavanje dinamike razvoja stanovništva i njegove migracije u kontekstu suvremenih društvenih i gospodarskih kretanja. Djelotvorno upravljanje i kontrola prometa doprinose povećanju sigurnosti, kvalitete prometne i transportno-logističke usluge, potiče ekonomski rast te smanjuje zagađenje zraka [?] i okoliša. Suvremeni društveno-ekonomski odnosi promijenili su prirodu kretanja ljudi i dobara zbog čega tradicionalni pristup postaje zastario i neprimjeren.

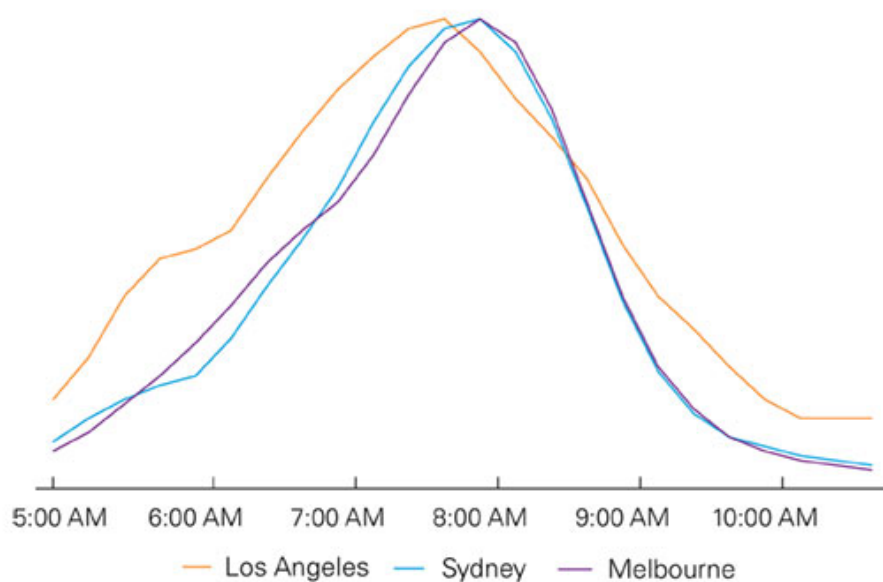
Tradicionalni modeli procjene pretpostavljaju da se širina vremenskog vrhunca u budućnosti neće mijenjati [?]. Širina vremenskog vrhunca već sada razlikuje se od grada do grada (Slike ??). Predviđanje promjena vrhunaca u bliskoj i dalekoj budućnosti zahtjeva odabir dobre vremenske rezolucije [?].

Pretpostavka da je referentna matrica *a priori* točna onemogućuje da se sagleda je li nova točnija i/ili kvalitetnija. Kako prikazati kvalitetu -e bez

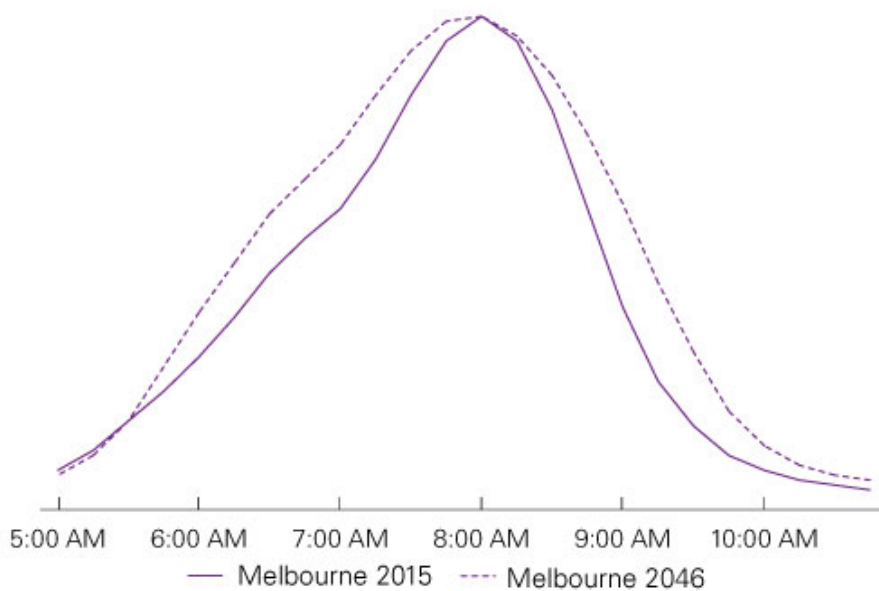
oslanjanja na istinitu referentnu POM-u koja porastom značajnijih i bržih promjena povećava utjecaj pristranosti kretanju u prošlosti? Postoji potreba za objektivnom procjenom kvalitete -e definiranjem odnosnih parametara kvalitete.

je kao izvor podjednako zastupljen u zemljama u razvoju i razvijenim zemljama s penetracijom oko 40%-50% stanovništva za pojedine operatore javne pokretne mreže [?]. **Praksa odbacivanja velikog dijela uzorka zbog niske potpunosti putanja može se zamijeniti rekonstrukcijom putanja uvažavajući njihov kontekst [?].**

Analiza kretanja stanovništva koja uzima u obzir kontekst proširuje područje primjene van analize prometa [?]. Mogućnost preciziranja odredišta s razine zone na razinu zgrade i određivanje svrhe (motiva) kretanja nije obuhvaćeno procjenom kvalitete -e.



(a) Širina jutarnjeg vrhunca prometne potražnje u 3 grada (oznake vremena su u lokalnom vremenu za svaki grad)



(b) Predviđena širina vrhunca prometne potražnje

Slika 4.1 Širina jutarnjeg vrhunca prometne potražnje - vremenski zavisno modeliranje [?]

4.2 Kontekstualizacija polazišno-odredišne matrice

Definicija 4.2.1. Putovanje kao informacijski objekt je entitet kojeg čine: **polazišni položaj**, **odredišni položaj**, **vremenska oznaka početka**, **vremenska oznaka kraja** i **kontekst** putovanja (ako je poznat).

Definicija 4.2.2. Kontekst putovanja je skup podataka koji definira polazište, odredište, **način prijevoza**, **svrhu** i **trajektoriju** putovanja.

Kontekst može biti izvučen iz podataka o kretanju ili iz vanjskih izvora podataka. Putovanja svake kontekstne kategorije čine po jednu kontekstualiziranu POM-u.

Kontekst specificira socio-ekonomske aktivnosti koje su pokretači putovanja, odnosno motivatori individualne i grupne mobilnosti. Preciznost određivanja konteksta definira izbor metode procjene -e.

Ako svrha putovanja nije izvorno navedena kao opisna veličina (prediktor) u ulaznim podacima za razvoj -e, moguće je o kontekstu putovanja zaključiti iz baza neizravnih kontekstualiziranih podataka trećih strana (npr. zapisi o korištenju telekomunikacijske mreže dopunjeni prostornim podacima) uz korištenje odgovarajućih postupaka prostorne statistike i analitike.

4.2.1 Primjeri korištenja konteksta u analizi kretanja

4.2.1.1 Studija A - 3 tipa privlačenja

Alhazzani i ostatak autora kvantizirali su vezu između 23 namjenske kategorije Točaka Interesa (engl. *Points of Interest*,) i onoga što nazivaju tipovi privlačenja. Namjenske kategorije definirane su iz službene gradske baze (npr. tvornice, bolnice, javne škole, religijski objekti, hoteli, knjižnice, sportski objekti). Namjenske kategorije mogu se podijeliti u tri skupine prema tipu privlačenja. Tipovi se razlikuju po gustoći ukupnog ulaznog toka, udaljenosti i disperziji polazišta. Primjerice, sveučilišta će privlačiti stanovnike sa svih strana grada (disperzija), podjednako iz udaljenih i bližih dijelova (udaljenost), dok će osnovne škole privlačiti samo iz bližih dijelova. Koriste podatke iz službene baze s 12,000 i zapisima iz razdoblja od mjesec dana [?]. Zone su okarakterizirane objektima koji se nalaze u njima.

4.2.1.2 Studija B - Model Raspodjele Toka

Inovativni pristup kontekstualizaciji toka predstavljen je u konferencijskom radu sa sudjelovanjem autorice ovog diplomskog rada [?]. Prostorni objekti podijeljeni su na osnovu pripadajućih opisnih podatak prema tipu socio-ekonomske aktivnosti uz koju su vezani (vidi dodatak ?? - OpenStreetMap). Definirano je 6 kategorija socio-ekonomskih aktivnosti: Dom (Home), Posao (Work), Zdravlje (Health), Edukacija (Education), Zabava (Leisure) i Ostalo (Other). **Razvijen je vjerojatnosni model koji ukupni ulazni tok u odredišnu zonu dijeli na 6 tokova usmjerenih prema objektima tih kategorija. Distribucija toka u kategorije ovisi o broju objekata pojedine kategorije u odredišnoj zoni i promatranom vremenskom okviru.** Jedna dijeli se na 6 kontekstualiziranih -a, po jednu za svaku od kategorija. -e dobivene iz javno dostupnih, anonimiziranih telekomunikacijskih zapisa na području kineskog grada Shenzhena korištene su u demonstraciji modela [?][?].

Definiranjem konteksta ulaznih tokova definiraju se nove informacije o kretanju unutar zona. Saznaje se o internom kretanju, čak i za metodu procjene koja bez kontekstualizacije nema točnost određivanja položaja veću od prostorne zone ().

4.2.1.3 Studija C - Planiranje održivih gradova

Modeli za procjenu potrošnje energije na razini zgrada temelje se na prostornom kapacitetu zgrada. Ispunjenost kapaciteta zgrade varira u vremenu što utječe na pogrešku u procjeni. Primjenom istog modela na razini grada pogreška raste s brojem zgrada. **Analizom kretanja stanovništva omogućuje procjenu ispunjenost kapaciteta zgrada i prosječne potrošnje energije.** Procjena kretanja iz zapisa prikladna je za analizu potrošnje energije na razini grada zbog velikog uzoraka stanovništva i neovisnosti o prometnoj infrastrukturi. Zgrade iz baze zgrada na području od interesa (grad Boston) podijeljene su u 3 tipa - stambene, industrijske i komercijalne. Model uzima u obzir ukupnu površinu zgrada i "radno vrijeme" zgrada koje nisu stambene. Zgrade unutar zone grupirane su u blokove. Ulazna putovanja dodijeljena su blokovima, a potom pojedinoj zgradi u bloku. Ovisno o pretpostavljenom odnosu između opterećenja i kapaciteta zgrada, u odnosu na standardne modele

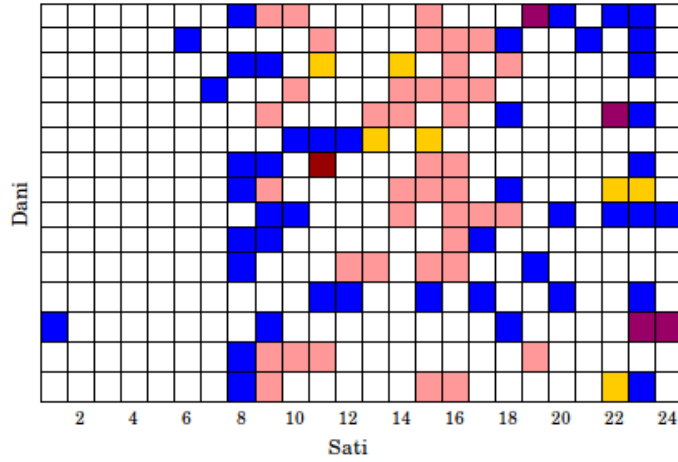
procjena potrošnje energije razlikuje se od +1% do - 15% za stambene te -4% do -21% za komercijalne zgrade [?] (standardni modeli procjenjuju potrošnju energije).

4.2.1.4 Studija D - Predviđanje odredišta u stvarnom vremenu

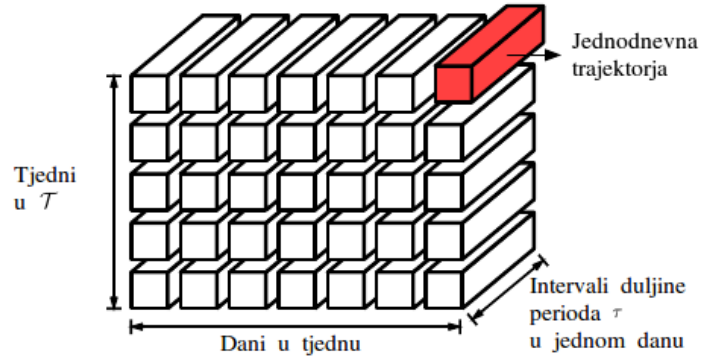
Studija se bavi predviđanjem odredišta započetog putovanja u pseudo-stvarnom vremenu na osnovu prepoznavanja svrhe putovanja. Svrha putovanja određuje se metodama strojnog učenja iz podataka o korištenju zemljišta () i poznatih uzoraka putovanja (model uči iz 5, 15 ili 30 putanja prethodnih putovanja istog pojedinca). Unaprijeđeni model predviđanja odredišta koristi svrhu putovanja kao kontekst u određivanju krajnjeg odredišta. Rezultati pokazuju značajno poboljšanje u točnosti i brzini predviđanja u odnosu na predviđanje bez uzimanja svrhe u obzir. [?]

4.2.1.5 Studija E - Rekonstrukcija rijetkih trajektorija

Rekonstrukcija rijetkih trajektorija predstavljena u radu [?] ima za cilj povećanje potpunosti (izraz ??) trajektorija definiranih iz . Temeljna ideja je iz poznatih položaja pojedinca zaključiti o nepoznatima (vidi sliku ??). je izvor visoke ne regularnost uzorkovanja i dugog razdoblja promatranja. Metodologija iskorištava redundantnost za oporavak podataka koji nedostaju. Redundantnost proizlazi iz regularnosti ljudskog kretanja, koje kreira uzorke posjećenih položaja kroz mnoge dane i tjedne razdoblja promatranja. Faktorizacija tenzora na kojoj se temelji metoda rekonstrukcije prikazana je na slici ??



Slika 4.2 Primjer trajektorije jednog korisnika, $\mathcal{T} = 15$ dana, $\tau = 1$ h. Svako polje predstavlja interval od 1 h, boje predstavljaju različite položaje, a bijela polja su položaji koji nedostaju. *Potpunost* = 0.27. Prema [?]



Slika 4.3 Primjer tenzora položaja jednog korisnika. Trajektorija je predstavljena kao kombinacije jednodnevnih trajektorija na dnevnoj i tjednoj bazi. Temelj rekonstrukcije je imati uvid u to - gdje je korisnik u periodu τ_n svaki dan u danom tjednu i gdje je korisnik u periodu τ_n tog dana u tjednu svaki tjedan kroz \mathcal{T} . Prema [?]

4.3 Teza

Postoji skup parametara kojima se objektivno mogu usporediti dvije na različiti način ili na isti način dobivene -e.

Rad nastoji tezu potkrijepiti opisom metodologije za vrednovanje matrica prema kvaliteti uvažavajući definirane odnosne parametre kvalitete.

4.3.1 Alternativni pristup vrednovanju polazišno-odredišne matrice

Točnost, rezolucija i obujam podataka ključni su parametri geografske informacije. Znanje o čemu parametri zavise je nužna baza za veće pothvate procjene vrijednosti geografskih informacija kao društvenog dobra [?].

Vrednovanje -e je proces koji ima za cilj procjenu obilježja kvalitete -e i znači identifikaciju obilježja kvalitete -e, njenog sadržaja, njene smislenosti, a izražava se sljedećim indikatorima: prostorno obuhvaćanje, rezolucija (prostorna, vremenska i tematska), zrnatost (prostorna, vremenska i tematska), putovanje kao parametar, gustoća informacija i širina toka.

4.4 Odnosni parametri kvalitete

4.4.1 Prostorno obuhvaćanje

Geoinformacijska znanost (geoinformatika) definira prostorno obuhvaćanje kao:

Definicija 4.4.1. Prostorno obuhvaćanje je geografsko područje gdje su prikupljani podatci, mjesto koje je predmet opažanja ili lokacija u fokusu aktivnosti. [?]

Definicija u ovom obliku odgovara definiciji prostornog obuhvaćanja kao parametra kvalitete -e. Prostorno obuhvaćanje kao kategorička mjera za kvalitetu -e podrazumijeva kategorije:

1. isključivo prometna infrastruktura (ceste, željeznica, javni prijevoz)

2. područje neovisno o prometnoj infrastrukturi (,)

Prostorno obuhvaćanje uvjetuje koji načini kretanja su obuhvaćeni -om.

150 od 175 zemljama članica Svjetske Zdravstvene Organizacije ima neki oblik zakona o zabrani korištenja mobilnih telefona tijekom uspravljanja vozilom [?]. Nije istraženo odražava li se to i kako na zastupljenost takvih putovanja u -i dobivenoj iz .

4.4.2 Rezolucija

Geoinformacijska znanost rezoluciju općenito definira kao količinu detalja u setu podataka [?]. Degbelo u radu "Pet općih obilježja rezolucije" [?] kao prvo obilježje navodi: **rezolucija (geografskih podataka) je pozitivno korelirana s točnošću i količinom podataka**. Što je veća razina detalja u prikazu, veća je sličnost tog prikaza 'istini' (savršenom prikazu). Što se više detalja pohranjuje, veća je količina podataka.

Isti rad napominje da *prisutnost rezolucije otvara pitanje "Koja je optimalna rezolucija i postoji li optimum?"*. Trud tek treba biti uložen u razvoj potpuno razrađene teorije optimalne **prostorne, vremenske i tematske rezolucije** koja će uzeti u obzir specifičnosti prikupljanja podataka (osjetila, ljudska osjetila, anketa, jesu li podaci proizašli iz prethodnih opažanja ili ne) i zadatak od interesa [?].

Vremenski nezavisna ima prostornu, vremenski zavisna prostornu i vremensku rezoluciju. Kontekstualizirana ima tematsku rezoluciju. Alternativni pristup vrednovanju -e definira prostornu i vremensku rezoluciju kao:

Definicija 4.4.2. Prostorna rezolucija je najmanja prijedena udaljenost koju je moguće identificirati.

Udaljenost je kontinuirana veličina, a radi pojednostavljenja metodologije definirana je kao kategorička s kategorijama:

1. $< 15m$
2. $15 - 50m$

3. $50 - 100m$
4. $100m - 500m$
5. $500m - 1km$
6. $1km - 2km$
7. $2km - 3km$
8. $> 3km$

Rezolucija nije svojstvo -e koliko je svojstvo izvora podataka i postupka procjene. Prostorna rezolucija -e iz CDR je najmanja udaljenost između središta 2 zone. (Pod niže navedenim ograničenjem ?? na putovanja iz istog izvora podataka - te dvije zone ne smiju biti susjedne).

Definicija 4.4.3. Vremenska rezolucija je najmanji vremenski odmak koji je moguće identificirati.

Vremenska rezolucija ovisi o periodu uzorkovanja položaja.

Geoinformacijska znanosti definira tematsku rezoluciju kao količinu kategoričkih detalja na *tematskoj karti*. Izražava se brojem tipova pokrova (engl. *land cover*) ili brojem tipova namjene zemljišta (engl. *land use*) [?]. Kontekstualizirane -e su procijenjene na temelju namjene zemljišta ili objekata, dakle koriste tematsku rezoluciju.

Definicija 4.4.4. Tematska rezolucija svrhe je broj kategorija svrhe putovanja definirane iz namjene krajnjeg odredišta.

Primjer kategorija i njihovih tematskih rezolucija:

- dom, posao (2)
- stambeni, industrijski, komercijalni (3) [?]
- dom, posao, edukacija, razonoda, zdravlje, ostalo (6) [?]
- tvornice, ambasade, sveučilišni objekti, bolnice, javne škole, ostali obrazovni objekti, stanice javnog prijevoza, knjižnice, klinike, religijski objekti, hoteli, sportski objekti, banke itd. (23) [?]

Razvrstavanjem putovanja u zasebne -e prema kategorijama dobije se broj -a jednak tematskoj rezoluciji.

Definicija 4.4.5. Tematska rezolucija načina kretanja je broj kategorija načina kretanja po kojim je moguće razvrstati putovanja.

Primjer kategorija i njihovih tematskih rezolucija:

- nadzemna željeznica (gradski javni prijevoz) (1) - izvor zapisi transakcija sustava Pametnih Kartica (engl. *Smart Card*) za prijevoz [?]
- taxi (1) - izvor zapisi GPS putanja taksi vozila
- osobna vozila, kamioni, javni prijevoz (3)
- auto, vlak, avion (3) - izvor CDR [?]
- pješaci, osobna vozila, vozila javnog prijevoza, kamioni (4)

4.4.3 Zrnatost

Zrnatost je stupanj do kojeg je sustav rastavljiv u manje dijelove, bilo sam sustav, njegov opis ili opažanje - stupanj do kojeg je veći entitet djeljiv [?].

Prostorna zrnatost kao odnosni parametar kvalitete izražava se kroz broj dijelova na koje je područje podijeljeno.

Definicija 4.4.6. Prostorna zrnatost -e je broj dijelova (zona) na području interesa. Iščitava se iz dimenzije -e.

Prostorna zrnatost ovisit će o prostornoj podjeli područja, gustoći baznih stanica, stupnju agregacije zona i prostornoj rezoluciji.

Definicija 4.4.7. Vremenska zrnatost je broj vremenskih okvira u jednom danu. Iščitava se iz broja -a procijenjenih za vremensko razdoblje $\mathcal{T} = 1$ dan.

Kategorije vremenske zrnatosti (vremenski okvir i pripadajuća zrnatost):

1. < 15 min (> 86)
2. 15 min (86) [?][?]
3. 20 min (72)

4. 30 min (48) [?]

5. 1 h (24) [?][?][?]

6. 2 h (12)

7. 3 h (8) [?][?]

8. 4 h (6)

9. 6 h (4)

10. 8 h (3)

11. 12 h (2)

12. 24 h (1) [?]

Masno otisnute kategorije zastupljene su u literaturi.

Najveća vremenska zrnatost koju je moguće postići procjenom -e ovisit će o vremenskoj rezoluciji u kojoj postoje podaci za procjenu.

Definicija 4.4.8. Tematska zrnatost je najmanja prostorno razlučiva jedinica koja ima tematsku rezoluciju.

1. zona

2. objekt [?] [?]

Objekt u zoni je najmanja prostorno razlučiva jedinica kod procjene matrica koja uključuje kontekstualizaciju toka na temelju objekata u odredišnoj zoni (studija B ??).

4.4.4 Putovanje kao parametar

Putovanja koja se protežu kroz više perioda u pravilu se dodjeljuju vremenskom okviru u kojem započinju ili u kojem završavaju. Autori [?] [?] [?] sva putovanja dodjeljuju intervalu u kojem je putovanje započeto, dok neki procjenjuje obje vrste -a [?]. Naredni odjeljci ne definiraju putovanje kao parametar, već samo ističu problematiku koja se javlja kod definicije putovanja općenito.

4.4.4.1 Ograničenja na putovanja iz izvora podataka -

Putovanja kod vremenski zavisnih -a kreiranih iz moguće je svrstati u nekoliko kategorija na osnovu prihvatanja ili odbacivanja određenih kretanja kao putovanja. Putovanjima se ne smije obuhvatiti lažno kretanje uzrokovano uravnoteženjem opterećenja i lažno zaustavljanje uzrokovano neuniformnom gustoćom zapisa pojedinaca u vremenu. Lažna kretanja i zaustavljanja u literaturi nastoje se izbjeći postavljanjem nekih od ograničenja:

1. putovanja u sve ćelije, bez internih putovanja (bez dijagonale)
2. putovanja u sve ćelije osim u susjedne (Slika ??)
3. agregacija susjednih ćelija - -e s internim putovanjima

Definicija 4.4.9. Interno putovanje je putovanje koje se odvija unutar jedne zone, odnosno čije se polazište i odredište nalaze unutar iste prostorne zone. Suprotno internom je eksterno putovanje.

4. putovanja ograničena minimalnom i/ili maksimalnom udaljenosti [?]
5. putovanja ograničena minimalnom duljinom trajanja zaustavljanja
6. putovanja ograničena minimalnom i/ili maksimalnom duljinom trajanja kretanja
7. putovanja ograničena bojem događaja koji potvrđuju zaustavljanje i maksimalnim vremenskim razmakom između dva uzastopna događaja

Studija provedena na podacima za Senegal [?] koristila je dvije definicije putovanja. Generirano je u prosjeku 0.7 putovanja dnevno po korisniku u jednom slučaju, a 0.06 u drugom. Ograničenja u prvom slučaju su bila: minimalna udaljenost 3 km, maksimalna 100 km (ograničenje ??). Ograničenja u drugom slučaju su bila: vremensko na 1 sat minimalnog trajanja (ograničenje ??) i prostorno na udaljenost od barem 5 udaljenosti baznih stanica (ograničenje ??). Kratka putovanja bila su najzastupljenija. Distribuciju možemo vidjeti na slici ?? **Ograničenja na putovanja uvjetuju njihov konačan broj odnosno širinu toka.**

4.4.4.2 Kontekstualizacija

Komutacijska putovanja (engl. *commuting trips*) odnosno putovanja na posao i putovanja kući značajna su po tome što su regularna - slijede dnevni ritam [?], što ih čini predvidivima [?]. Goulding u priručniku za kreiranje -e iz ističe da je važno definirati što se smatra putovanjem - zastane li netko na benzinskoj crpki na putu s posla kući, je li riječ o jednom putovanju ili dva [?]. Prihvatanjem svakog zaustavljanja kao kraja (primjerice ukidanjem ograničenja ??), u -u se uvrštavaju podputovanja. Uvrštavanjem podputovanja dobiva se tranzitna (spomenuta u ??). Tranzitna uopće neće opisivati kretanja između polazišta i odredišta *u suštini* stoga treba izbjegavati usputna zaustavljanja [?].

Prema duljini zaustavljanja nastoji se odrediti je li neko zaustavljanje doista kraj putovanja [?]. Poznavanje konteksta putovanja, npr. objekata u blizini zaustavljanja, omogućuje da se kraj putovanja ne temelji isključivo na osnovu duljine trajanja zaustavljanja, već da se u obzir uzme svrha zaustavljanja.

Prednost suvremenih izvora podataka s višednevnim ili višemjesečnim razdobljem promatranja je uočavanje obrazaca i ponovljivost putovanja na dnevnoj ili tjednoj bazi [?]. Prepoznavanje obrasca putovanja pojedinca u tom razdoblju omogućuje rekonstrukciju putovanja u periodima kada podaci nedostaju (sekcija ??). Chen putovanja pojedinca u prošlosti i budućnosti od promatranog putovanja naziva upravo *kontekstom* u metodologiji opisanoj u radu [?].

Mjera koja bi govorila o kvaliteti odabrane definicije putovanja prema razmotrenoj problematici mogla bi biti omjer tranzitnih (podputovanja) i ne-tranzitnih putovanja (putovanja sa svrhom) zabilježenih u -i. Kako izostaje konvencionalni dogovor oko definicije putovanja i tranzicije (podputovanja) ova mjera nije obuhvaćena ovim radom.

4.4.5 Gustoća informacija

Definicija 4.4.10. Gustoća informacija je omjer ispunjenih ćelija i Nul-ćelija u -i.

-e dobivene iz anketa u kućanstvima i iz rijetko su popunjene, odnosno većinski su ispunjene Nul-ćelijama (eng. *Null-cells*, *Zero-cells*) [?]. -e kreirane iz GPS ili

zapisa gusto su ispunjene, najčešće vrlo malim vrijednostima umjesto nula [?]. -e izrađene iz uzoraka nakon procjene se skaliraju da bi obuhvatile cijelu populaciju. Postupci skaliranja nisu obuhvaćeni ovim radom. Skaliranje uključuje množenje -e određenim faktorom, a u tom slučaju Nul-ćelija zapravo ne mijenja vrijednost.

Dva su uzroka praznim ćelijama, nepostojanje putovanja između tih parova zona ili su postupkom procjene putovanja nezabilježena.

4.4.6 Ukupna širina toka

Definicija 4.4.11. Ukupna širina toka je ukupan broj odlazaka/dolazaka po vremenskom okviru za cijelu matricu - suma svih elemenata -e. (vidi sliku ?? F_{ij}).

Svi dosad spomenuti parametri utjecat će na ukupnu širinu toka. Osim njih, utjecat će veličina i reprezentativnost uzorka stanovništva te k-anonimizacija.

Calabrese [?] uzima samo slučajno odabranih 25% korisnika (1 milijun) da pojednostavi analiza. Graells-Garrido [?] uzima slučajno odabranih 100,000 korisnika jer je referentna procijenjena iz ankete provedene na uzorku iste veličine. Uzimanje samo najaktivnijih korisnika (onih s najviše zapisa) zastupljeno je u istraživanjima procjene -e iz [?] [?] [?]. Chen analizira razloge odbacivanja i predlaže rješenje. Rješenje je metoda rekonstrukcije putanja koja osigurava zadržavanje velikog broja (manje aktivnih) korisnika (sekcija ??).

Slabo obuhvaćanje krajnjih dobnih skupina (najstarije i najmlađe dobne skupine) u procjeni -a iz zbog slabe zastupljenosti mobilnih uređaja ne tumači se kao pristranost, uz napomenu da su to ujedno skupine sa generalno slabom mobilnosti u odnosu na ostale [?]. Goulding smatra da se eventualna pristranost može zanemariti kod tržišta gdje veliki postotak stanovništva (npr. 92%) posjeduje mobilni uređaj [?].

4.4.7 K-anonimizacija

K-anonimnost (engl. *k-anonymity*) određuje najmanju dozvoljenu širinu toka ili najmanju dozvoljenu populaciju zone. Širina toka manja od dozvoljene se zanema-

ruje (postaje Nul-ćelija). Razlog je smanjenje rizika od mogućnosti identificiranja pojedinaca, pogotovo u vrlo rijetko naseljenim područjima [?]. K-anonimizacijom treba osigurati:

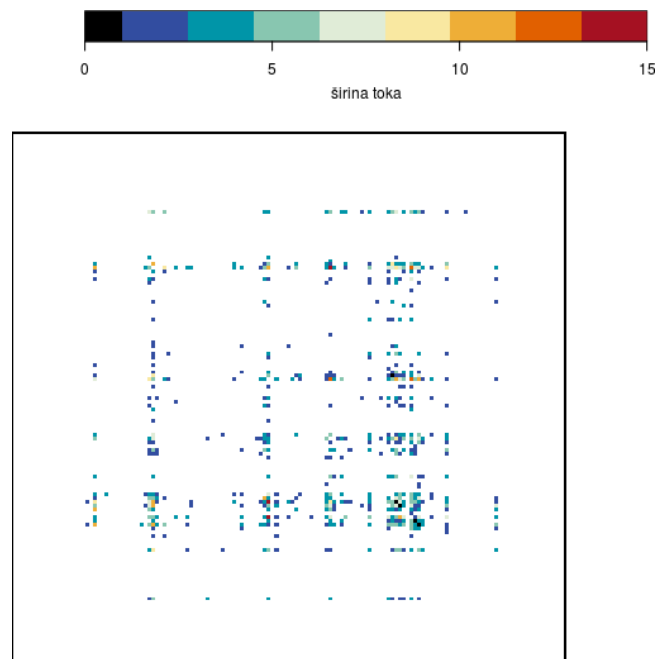
$$f_{ij} > k \times m, \forall i, j \quad (4.1)$$

gdje je:

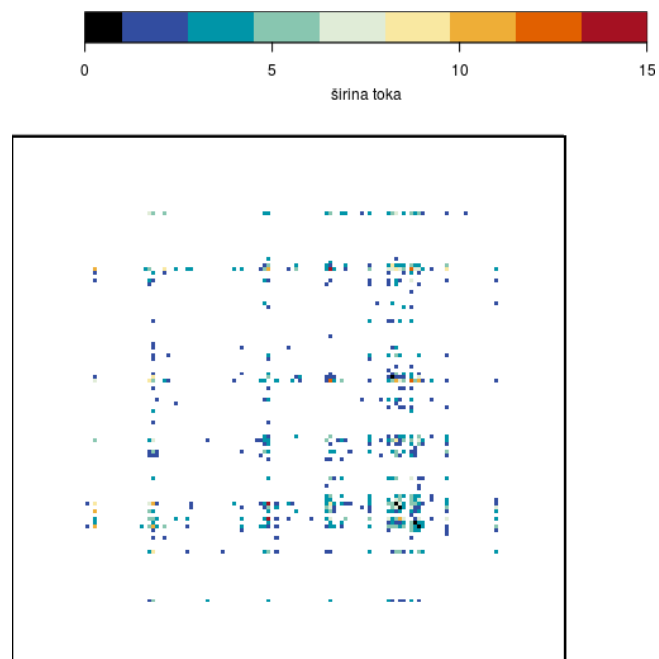
m - maksimalni broj putovanja za svakog pojedinca

k - odabrana vrijednost kojom se želi osigurava da na temelju širine toka pojedinac ne bude razaznat između $k - 1$ drugih pojedinaca.

Alternativno, ako su dostupni, iz podataka o populaciji kod definiranja zona k-anonimizacijom dovoljno je osigurati da je populacija u svakoj zoni veća od k [?].

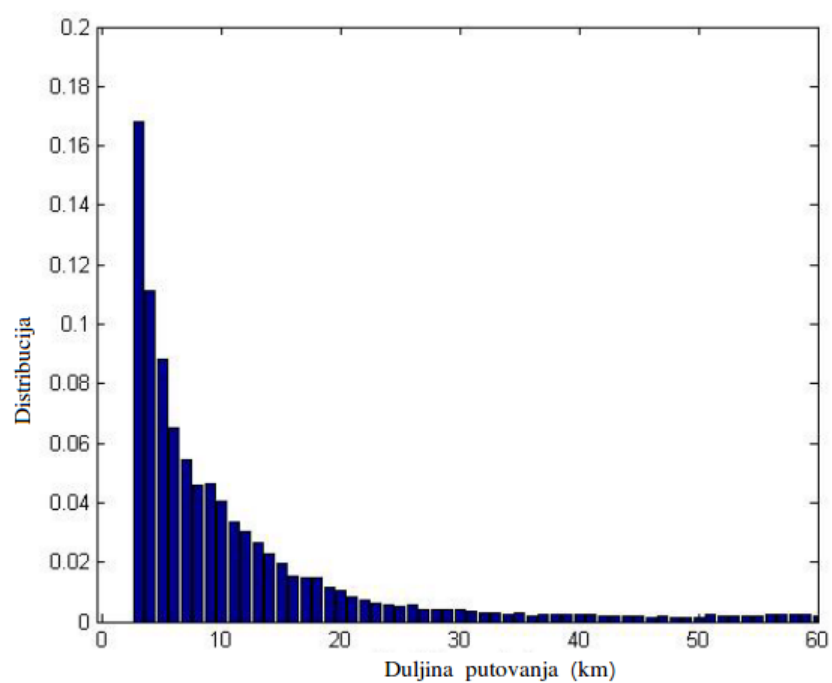


(a) POM s putovanjima u susjedne ćelije. Ukupna širina toka 2587

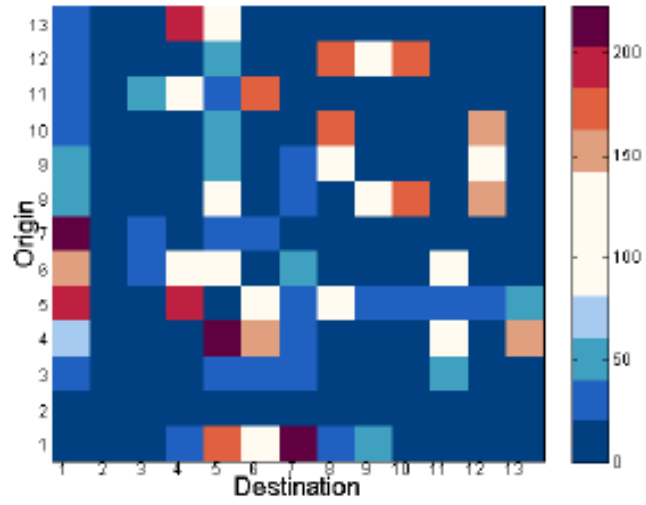


(b) POM bez putovanja u susjedne ćelije. Ukupna širina toka 2392

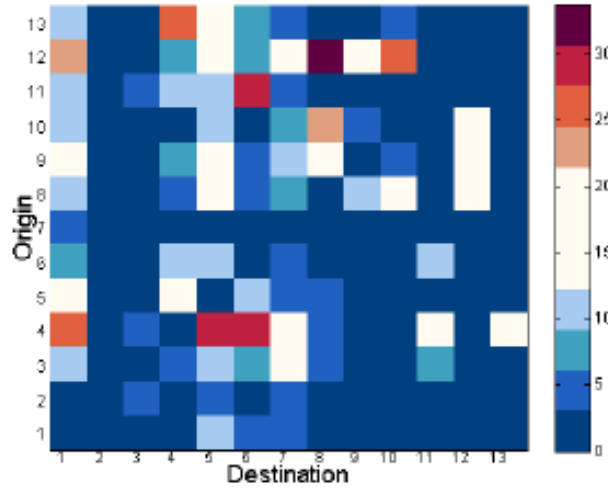
Slika 4.4 Primjer POM-e s i bez putovanja u susjedne ćelije.



Slika 4.5 Distribucija udaljenosti putovanja u kilometrima, Senegal 2016. [?]. U studiji su izbačena putovanja kraća od $3km$ (prostorna rezolucija $> 3km$)



(a) POM u kojoj je putovanje definirano kao zabilježeni pomak između 2 antene - *kratka putovanja*



(b) POM u kojoj je putovanje definirano kao promjena položaja u odnosu na položaj u prethodnom periodu. Položaj u periodu je onaj na kojem korisnik pretežito boravi. Za jedan period se pretpostavlja da je korisnik na prebivalištu, za drugi da je na radnom mjestu - matricu čine putovanja od prebivališta na posao (engl. *commuting trips*)

Slika 4.6 Dvije POM-e - produkti različitih definicija putovanja. Dobivene su iz istog izvora zapisa [?]

Poglavlje 5

Metodologija

5.1 Pretpostavke usporedbe

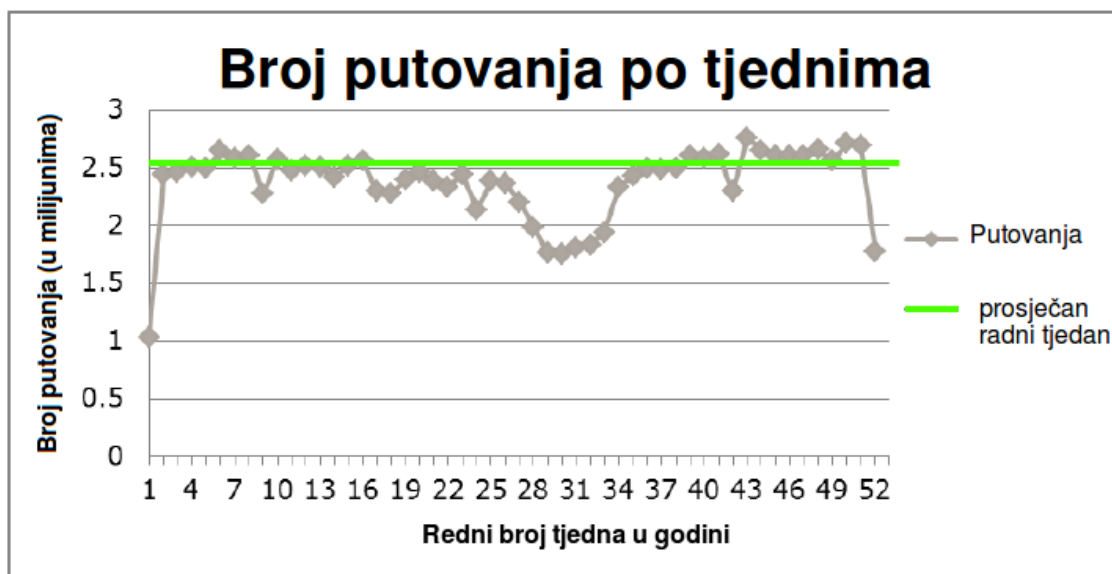
-e koje se uspoređuju trebaju ispunjavati pretpostavke usporedbe, odnosno imati sljedeća zajednička obilježja:

- isto područje/grad/država
- isto doba godine

Zabilježenu varijaciju po tjednima u jednoj godini iz istraživanja procjene -e iz podataka o javnom prijevozu na području Nizozemske možemo vidjeti na slici ???. Prednost novih postupaka procjene -e (npr. iz) nad procjenama iz anketa stanovništva je da mogu uhvatiti tjedne i sezonske uzorke [?].

- ako se radi o jednom danu - idealno isti dan u tjednu

Rezultati nacionalne ankete u Sjedinjenim Američkim Državama pokazuju da je prosječan broj dnevnih putovanja po stanovniku 4.18 radnog, te 3.86 neradnog dana, što je u skladu s zaključcima suvremenih postupaka procjene [?][?]. Osim očekivane razlike između radnog i neradnog dana, uočeno je da je petak (zadnji radni dan u tjednu), drugačiji od ostalih radnih dana [?][?][?]. Uočena je pojačana mobilnost petkom na razini regije (Boston Metropolitan Area) [?] i na razini države (Obale Bjelokosti) s povećanjem od 35% u odnosu na nedjelju [?] na temelju -a izrađenih iz . Uočena je smanjena mobilnost vezana za javni



Slika 5.1 Broj putovanja javnim prijevozom zabilježenih *OV-chipkaart* pametnim karticama u Nizozemskoj po tjednima u jednoj godini. Zelena linija označava prosječnu vrijednost broja putovanja radnog tjedna [?]

prijevoz (izvor za procjenu -a *Pametne kartice za javni prijevoz*) na razini regije (Southeast Queensland, Australija) [?]. Petak ima i nešto drugačiji dnevni uzorak po satima (pojačanje prometa prema kraju radnog vremena započinje ranije) [?]. Državni praznici, elementarne nepogode i druge izvanredne i predvidljive situacije unose odstupanja od ustaljenih uzorka kretanja. Privremene turističke atrakcije i sportski događaji također mogu utjecati na odstupanje od uobičajenog uzorka kretanja [?].

- ista definicija dana

Dio putovanja započinje u jednom, a završava u drugom danu. Autori [?][?] zastupaju definiranje početka/završetka dana u 3:00 idućeg dana, umjesto u 0:00, kako dio putovanja ne bi bio izostavljen zato jer se proteže kroz 2 dana.

- ista kategorija putovanja (tranzitna, kontekstualizirana, *komutacijska*)
- obje -e napravljene objektivno na reprezentativnom uzorku

5.2 Usporedba prema odnosnim parametrima kvalitete

Ako se ustanovi da -e ispunjavaju pretpostavke usporedbe, uspoređuje ih se prema parametrima definiranim u ???. Metodologija definira da vrijedi sljedeće:

- prostorno obuhvaćanje: **je bolja ako je prostorno obuhvaćanje veće** (područje neovisno o prometnoj infrastrukturi > područje ovisno o prometnoj infrastrukturi)
- gustoća informacija: **je bolja što je gustoća informacija veća (što je manji broj Nul-ćelija)**
- ukupna širina toka: **je bolja što je ukupna širina toka veća**
Podrazumijeva se isti vremenski okvir npr. od 6:00 do 9:00, cijeli jedan dan i dr.
- zrnatost: **je bolja što je zrnatost veća, pod uvjetom da je dovoljno mala da osigurava k-anonimnost**
K-anonimnost smanjuje rizik da identitet pojedinca bude razaznat od $k - 1$ drugih pojedinac (??). Povećanjem bilo koje zrnatosti tokovi postaju sve uži. Uvažavanje k-anonimnosti postavlja ograničenje na zrnatost [?]. Tradicionalni postupci procjene (podatak za Ujedinjeno Kraljevstvo) dozvoljavaju širinu jednog toka između 12 i 15. Agencija za analizu kretanja iz Teralytics [?] koristi k između 5 i 10.
- rezolucija: Sekcija ?? govori o pozitivnoj korelaciji rezolucije i točnosti. Odbacivanjem točnost kao jedinog pokazatelja kvalitete pitanje postaje koja rezolucija je optimalna za prikaz dnevnih urbanih migracija u -i. Visoka rezolucija podataka omogućuje visoku zrnatost. Primjerice, ako možemo odrediti završetak putovanja u minutu, vremenski okvir teoretski može biti $< 15min$. Odabir niske zrnatosti umanjuje potrebu za visokom rezolucijom. Primjerice, ako je vremenski okvir širine 12 sati (2 perioda u danu) nema potrebe da je početak i završetak putovanja određen u minutu.

Optimalne vrijednosti parametara naposljetku će ovisiti o području primjene -e.

Poglavlje 5. Metodologija

Određivanje optimalnih vrijednosti parametara zahtjeva dodatna istraživanja koja bi se nadovezala na metodologiju predstavljenu u ovom radu. Primjerice, rezolucija koja dobro odgovara u jednom može biti nepotrebno visoka u drugom području primjene. Odabir niže rezolucije smanjuje količinu pohranjenih podataka [?] i može smanjiti vrijeme izvođenja postupka procjene.

5.2.1 Postupak odlučivanja

Metodologija definira svaki parametar kao jednako vrijedan u donošenju konačne odluke o kvaliteti matrice. POM koja je bolja u više parametara ukupno je kvalitetnija. Parametar koji se ne odnosi ni na jednu matricu se ne uzima u obzir.

Poglavlje 6

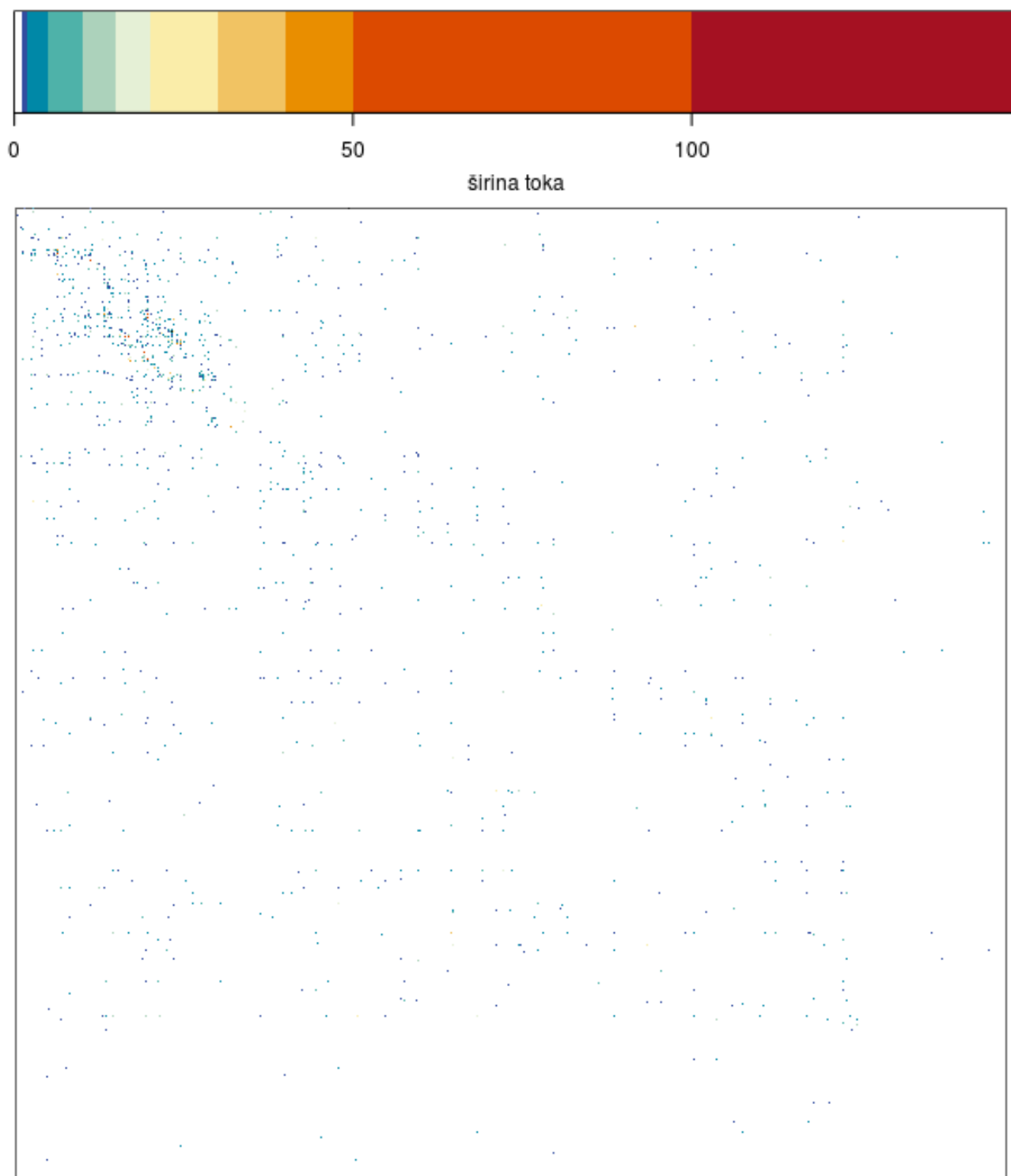
Rezultati

Metodologija je demonstrirana na slučaju usporedbe dvije -e dobivene s podacima iz različitih izvora i različitim postupcima procjene. Duljina razdoblja prikupljanja podataka za obje -e je $\mathcal{T} = 24h$, na isti datumu. Za implementaciju demonstracije metodologije korišteno je programsko okruženje za statističko računarstvo R.

A je dobivena iz anonimiziranih javno dostupnih telekomunikacijskih zapisa za grad Shenzhen () [?] postupkom procjene opisanim u radovima [?] [?] [?]. Rezultat procjene je 8 -a za vremenske okvire od 3 sata i jedna za cijeli dan. (Vidi sliku ??)

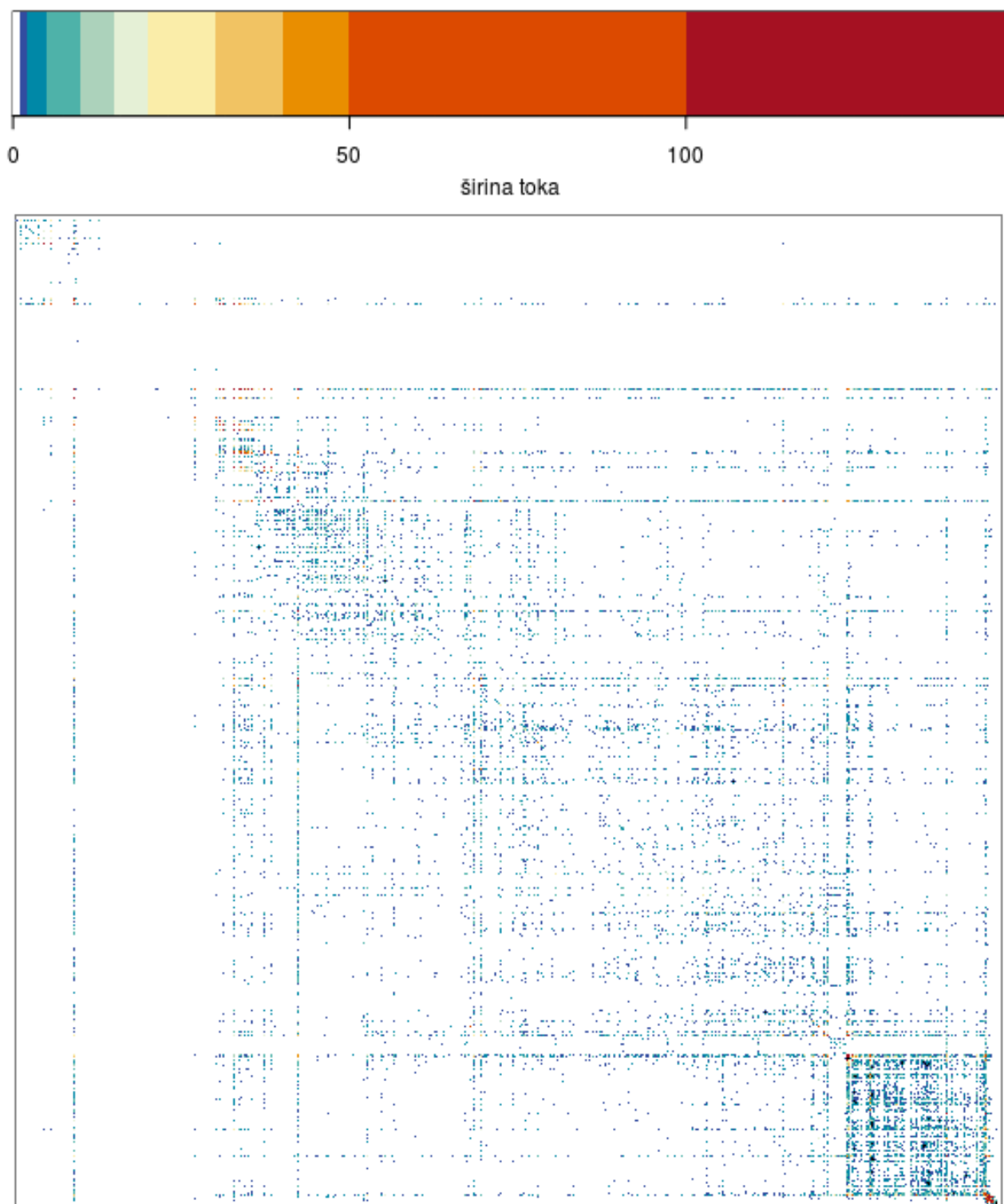
B je dobivena iz javno dostupnih zapisa položaja i statusa (ima putnika/ nema putnika) Taxi vozila u gradu Shenzhenu [?] postupkom procjene opisanim u dodatku ???. Rezultat procjene je 8 -a za vremenske okvire od 3 sata i jedna za cijeli dan. (Vidi sliku ??)

Poglavlje 6. Rezultati



Slika 6.1 POM A

Poglavlje 6. Rezultati



Slika 6.2 POM B

Poglavlje 6. Rezultati

Tablica 6.1 Tablica usporedbe

POM	prostorno obuhvaćanje	gustoća informacija	prostorna zrnatost	vremenska zrnatost	tematska rezolucija načina kretanja	ukupna širina toka
A	neovisno o prometnoj infrastrukturi	0.008830772	1090×1090	8	0	30878
B	cesta	0.08683223	1090×1090	8	1	427646

POM	vremenska rezolucija	prostorna rezolucija
A	*	$\approx 75m$ *
B	5s	$\approx 10m$ (GNSS)

* vremenska rezolucija je neodređena,
prostorna rezolucija za općenito, prema preporukama [?]

Prema odnosnim parametrima kvalitete definiranim u sekciji ?? i metodologiji usporedbe definiranoj u poglavlju ??, B ukupno ostvaruje bolji rezultat.

Poglavlje 7

Diskusija

Ovaj rad uvodi novi pogled na usporedbu i vrednovanje -a dobivenih iz različitih izvora i različitim postupcima procjene temeljen na vlastitim definicijama parametara kvalitete. Osim parametara kvalitete, rad definira pretpostavke usporedbe dvije -e te postupak odlučivanja o njihovoj odnosnoj kvaliteti. Rezultat istraživanja je okvirna metodologija za usporedbu kontekstualiziranih -a. Rezultat istraživanja podržavaju početnu tezu.

Rezultat demonstracije metodologije je tablica ?? u kojoj je -a B ukupno ostvarila bolji rezultat odnosno kvalitetu. Metodologija nudi rješenje za objektivno vrednovanje matrica, temeljeno na parametrima kvalitete. Uzmemo li A kao referentnu matricu, a B kao onu koju se vrednuje, B ostvaruje bolji rezultat od referentne zbog boljih vrijednosti na području tematske rezolucije, ukupne širine toka i gustoće informacija.

Okvirna metodologija koja je predstavljena i demonstrirana u ovom radu u postupku odlučivanja definira sve parametre kvalitete kao jednako vrijedne, što ima veliki utjecaj na konačne rezultate usporedbe. Drugačije postavljeni odnosi parametara, primjerice dodjeljivanjem različite *težine* svakom parametru utjecali bi na konačni rezultat vrednovanja. Procijenjeni nedostaci metodologije zahtijevaju dublju analizu svakog definiranog parametra kvalitete i kriterija usporedbe zasebno, kao i njihovih međusobnih ovisnost. Cilj daljnjih istraživanja treba biti i određivanje optimalnih vrijednosti parametara te analiza -a svedenih s apsolutnih na relativne

vrijednosti te razmatranje apsolutne i relativne točnosti -e:

$$X \rightarrow Z = \frac{x - \bar{x}}{\sigma} \quad (7.1)$$

gdje je Z transformirana X .

Buduća istraživanja mogu ovom metodom ispitati ideju različite kvalitete u pojedinim dijelovima -e u vremenskoj i prostornoj domeni i sagledati neuniformnost vremenske i prostorne rezolucije podataka i procijeniti njen utjecaj na kvalitetu dijelova . Putovanja se u praksi dodjeljuju vremenskom okviru u kojem započinju ili završavaju bez obzira što se protežu kroz više perioda. *Klizeći prozor* zamišljen je kao pomični vremenski okvir koji obuhvaća samo putovanja koja mu u cijelosti pripadaju tj. započinju i završavaju unutar vremenskog okvira. *Klizeći* iz razloga što razdoblje promatranja ne mora biti podijeljeno u određen broj vremenskih okvira već se vremenski okvir može postaviti bilo kad. Prijedlog za danja istraživanja je i analiza kako i koliko *klizeći prozori* mogu utjecati na kvalitetu matrice te koji su izvori podataka pogodni za primjenu *klizećih prozora*. Potrebno ispitati utjecaj vrijednosti k u k -anonimizaciji na kvalitetu -e.

Poglavlje 8

Zaključak

Cilj rad bilo je definirati skup parametara kojima se objektivno mogu usporediti dvije polazišno-određene matrice, osmisliti i demonstrirati metodologiju vrednovanja matrica prema kvaliteti uvažavajući definirane odnosne parametre kvalitete.

Napravljen je pregled prethodnih istraživanja kojim je utvrđeno da se postupci vrednovanja -e temelje na sličnosti -e s *a priori* točnom (istinitom) referentnom -om. Točnost, odnosno sličnost nove -e s postojećom -om uzima se kao jedini parametar kvalitete nove -e. Postupci vrednovanja nisu usmjereni na objektivno vrednovanje. Objektivno vrednovanje trebalo bi omogućiti da koju se vrednuje eventualno pokaže bolju kvalitetu od referentne, istinite -e. Doprinos nove -e i postupka procjene na području određivanja konteksta putovanja, prostorne i vremenske rezolucije, područja obuhvaćanja slabo je mjerljiv postojećim postupcima vrednovanja.

Ovaj rad definira odnosne parametre kvalitete: prostorno obuhvaćanje, rezolucija (prostorna, vremenska i tematska), zrnatost (prostorna, vremenska i tematska), putovanje kao parametar, gustoća informacija i širina toka. Definira novu metodologiju vrednovanja -e usporedbom s referentnom -om po parametrima kvalitete. Pri usporedbi dviju -a metodologija definira svaki od parametara kao jednako vrijedan u procjeni ukupne kvalitete. Bolja je ona koja u više parametara pokazuje bolju vrijednost. Ovim postupkom vrednovanja -e kvalitetom je obuhvaćen kontekst putovanja, područje pokrivanja -e, popunjenost -e (broj nepraznih ćelija), njena ukupna širina toka, zrnatost i rezolucija.

Poglavlje 8. Zaključak

Nedostatak metodologije je taj što su parametri kvalitete definirani općenito pa definicija može zahtijevati prilagodbu postupku procjene. Svaki definirani parametar zahtjeva u budućnosti dublju analizu s ciljem definiranja njegovih optimalnih vrijednosti. Optimalne vrijednosti parametara trebaju uvažavati područje primjene -e i osiguranje anonimnosti pojedinaca čiji su podaci korišteni za procjenu -e. Potrebno je ispitati utjecaj jednake vrijednost svakog parametra u određivanju ukupne kvalitete i po potrebi definirati težinske faktore za parametre.

Parametri i metodologija definirani u ovom radu okvirno su usmjerenje k daljnjim istraživanjima na području vrednovanja i procjene .

Sažetak

POM (POM) je alat koji omogućuje opis i sustavnu statističku procjenu migracija stanovništva na nekom području u zadanom prostorno-vremenskom okviru. Polazišno-odredišna matrica (POM)-a služi za opis grupne mobilnosti i mjerenje socio-ekonomske aktivnosti u nekoj regiji, a najčešće se koristi u prometnoj znanosti za analizu i strateško planiranje prometnog opterećenja i prometne infrastrukture. Postoje brojni tradicionalni pristupi procjeni i vrednovanju POM-a, a dolaskom pasivno prikupljenih suvremenih izvora podataka o kretanju (zapisi usluga baziranih na lokaciji) omogućen je razvoj suvremenih pristupa i proširenje područja primjene. Postupci procjene koji uključuju kontekstualizaciju dnevnih migracija daju uvid u motive odnosno svrhu kretanja stanovništva. Postojeće metode vrednovanja POM-a definiraju vrijednost nove POM usporedbom i razmatranjem sličnosti s postojećom za isto područje. Postoji potreba da se definira i kvantizira kvaliteta POM-e kroz objektivne parametre. Ovaj rad predstavlja alternativni pristup vrednovanju POM-a, definira objektivne parametre kvalitete, pretpostavke usporedbe i postupak odlučivanja o kvaliteti POM. Definirana metodologija obuhvaća kontekstualizaciju kao dio kvalitete POM. Metodologija je demonstrirana na usporedbi dvije POM dobivene iz različitih izvora različitim postupcima procjene.

Ključne riječi — Polazišno-odredišna matrica (Izvorišno-odredišna matrica), kontekstualizacija, parametri kvalitete, usporedba, vrednovanje

Abstract

Origin-Destination Matrix (ODM) is a tool that enables description and systematic statistical estimation of human migrations in an area and within a give space-time frame. ODM is used to describe group mobility and measure socio-economic activity in a region, and is most commonly used in transport science for the analysis and strategic planning of transport load and transport infrastructure. There are a number of traditional approaches to ODM estimation and evaluation. The rise of passively collected location data (from location based services) has enabled development of new, modern approaches and broadens the field of use. Approaches that

involve contextualisation of daily migrations provide insight into the motives and te purpose of urban movements. The existing ODM valuation methods define the value of a new ODM by comparing it to an existing ODM and considering the simmilarity. There is a need to define and quantify the quality of ODM through objective parameters. This thesis presents an alternative approach to ODM evaluation, defines objective quality parameters, comparison conditions and decision process. Defined methodology captures contextualization as part of ODM quality. The methodololgy is demonstrated with comparing two ODMs obtained from different sources and with different estimation procedures.

Keywords —Origin-Destination Matrix (Trip Table), contextualization, quality parameters, comparison, evaluation

Dodatak A

Metrike za vrednovanje polazišno-odredišne matrice

A.0.1 Metrike za procjenu sličnosti polazišno-odredišne matrice s referentnom

Za procjenu kvalitete POM-a dobivenih isključivo anketiranjem u radu [?] korištena je mjera srednja apsolutna postotna pogreška (engl. Mean Absolute Percentage Error, MAPE), te je prikazano da se zadovoljavajuća razina kvalitete takvih POM-a postiže ako uzorak obuhvaća 50% populacije. Autor Cools ističe važnost korištenja drugih izvora uz ankete za izradu POM-a.

U radu [?] navedene su statističke mjere relativna pogreška (engl. Relative Error, RE), devijacija ukupne potražnje (engl. Total Demand Deviation, TDD), srednja apsolutna pogreška (engl. Mean Absolute Error, MAE), korijen iz srednje kvadratne pogreške (engl. Root Mean Square Error, RMSE) te najveća moguća relativna pogreška (engl. Maximum Possible Relative Error, MPRE) i razina prometne potražnje (engl. Travel Demand Scale, TDS) koji procjenjuju kvalitetu neovisno o referentnoj matrici (no MPRE ne dopušta pogreške u prebrojavanju prometa, dok TDS ovisi o topologiji mreže i odabiru ruta) [?].

U [?] korišten je *Pearsonov koeficijent korelacije* - r da bi se utvrdila **sličnost svakog retka POM-e** dobivene iz CDR s **retkom referentne** (ukupni izlazni tok

iz svake polazišne ćelije). Isti postupak korišten je za kontekstualizirane Home-Work (HW) i Work-Home (WH) POM-e dobivene iz Call Data Records ili Charging Data Records (CDR) u usporedbi s referentnim POM-ma dobivenim anketiranjem.

U svome radu [?] navodi se nekoliko mjera - R^2 , Geoffrey E. Havers statistika (GEH), korijen iz postotak srednje kvadratne pogreške (engl. Root Mean Squared Error percentage, %RMSE), uvodi novu mjeru Eigenvalue-Based Measure (EBM) (temeljenu na svojstvenim vrijednostima matrica) i procjenjuje pouzdanost POM-e dobivene iz sustava automatskog prikupljanja podataka u javnom prijevozu (autobus, vlak i trajekt). Spominje i Wasserstein metriku, mjeru koja ne uspoređuje samo vrijednosti parova istih ćelija (*elementwise*).

Spearmanov koeficijent korelacije ranga korišten je u [?] za procjenu sličnosti POM-a dobivenih iz CDR sa tada aktualnim POM-ma dobivenim anketiranjem.

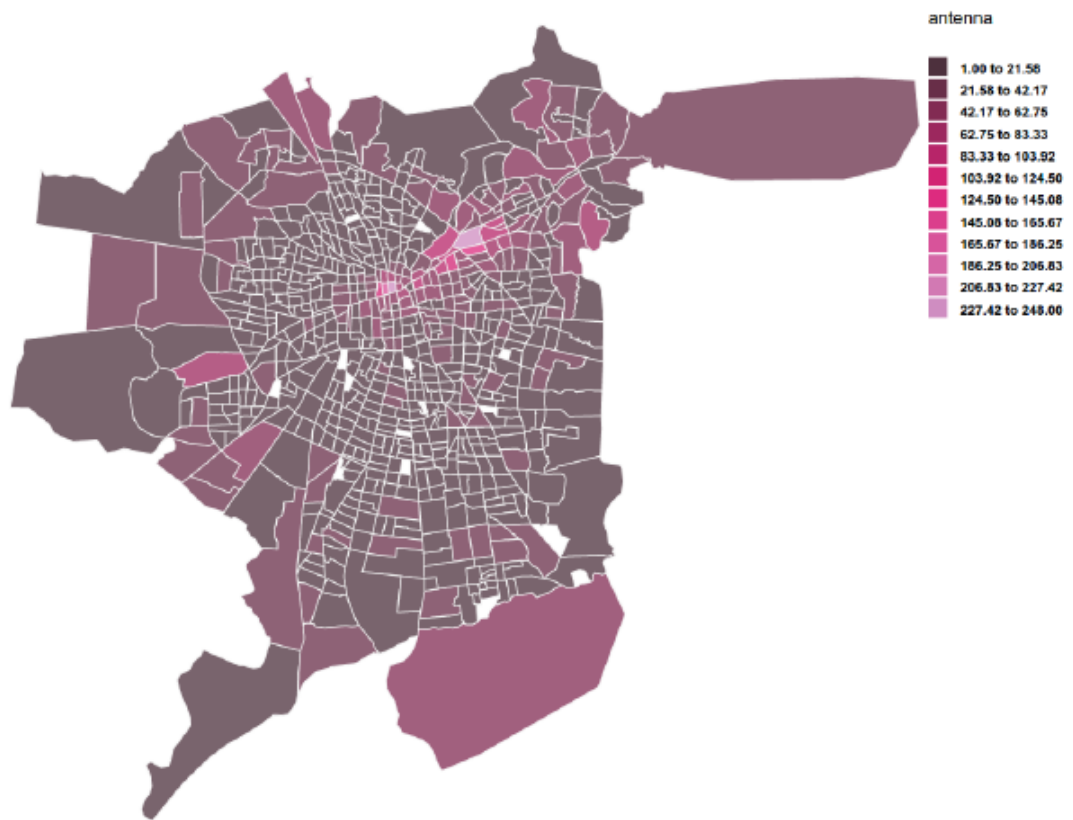
Dodatak B

Prilagodba prostorne podjele

Definiramo li točnost matrice isključivo na temelju sličnosti svakog njenog elementa s ekvivalentnim elementom u *grand truth* matrici nužno je osigurati da prostorna podjela matrice odgovara prostornoj podjeli referentne ili je potrebno obje matrice svesti na zajedničku prostornu podjelu.

B.1 Objedinjenje baznih stanica

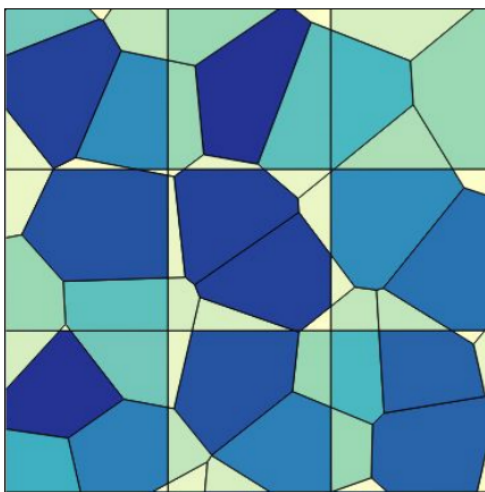
Pregledom literature utvrđeno je da se koriste *k-sredina* i *mean-shift* algoritam za grupiranje baznih stanica u broj grupa jednak broju Traffic Analysis Zone - prometna zona (TAZ) na području interesa [?] [?] [?] ili pak jedinicama samouprave (eng. *municipalities*) kako bi POM-e u konačnici bile usporedive s postojećima, dobivenim iz anketa [?] [?]. Nije utvrđivano kolika se greška uvodi ovim postupcima. Primjer gustoće tornjeva baznih stanica na područjima anketnih jedinica prikazan je na slici ??



Slika B.1 Prostorna podjela na anketne jedinice u gradu Santiagu. Gustoća tornjeva baznih stanica po anketnim jedinicama. Gustoća varira od 1 do 250 tornjeva po anketnoj jedinici. [?]

B.2 Interpolacija

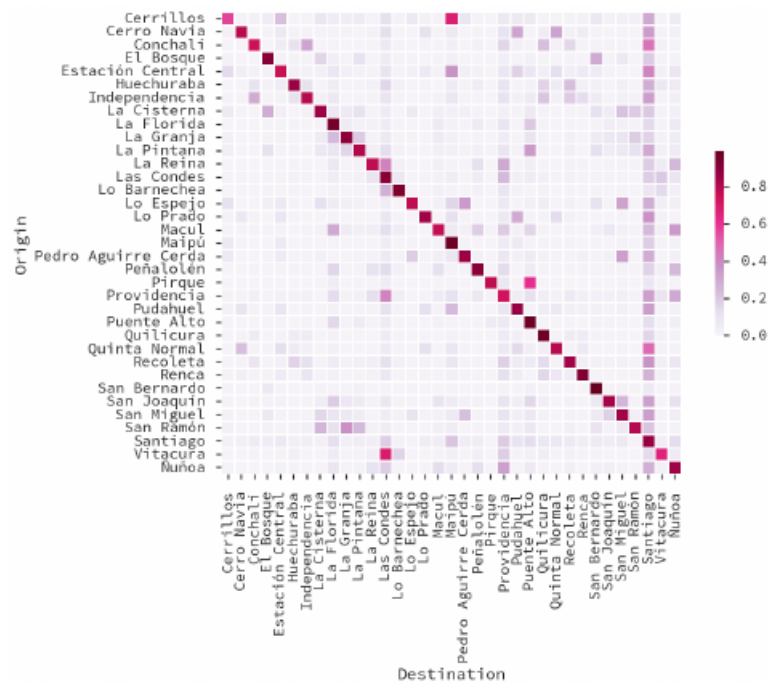
Postupak konverzije POM-a iz prostorne podjele na Voronoi ćelije u drugu prostornu podjelu, npr. uniformnu kvadratnu mrežu (ćelije $1km^2$), opisan u priručniku [?]. U postupku određivanja postotka ukupnog toka Voronoi ćelije koji će se dodijeliti novoj kvadratnoj ćeliji predlaže uzeti u obzir: površinu preklapanja tih ćelija, broj zgrada ili ukupnu površinu zgrada (uključujući katove) na području preklapanja tih ćelija. Na slici ?? je prikazan treći oblik interpolacije gdje su korišteni podaci o ukupnoj kvadratnoj površini zgrada.



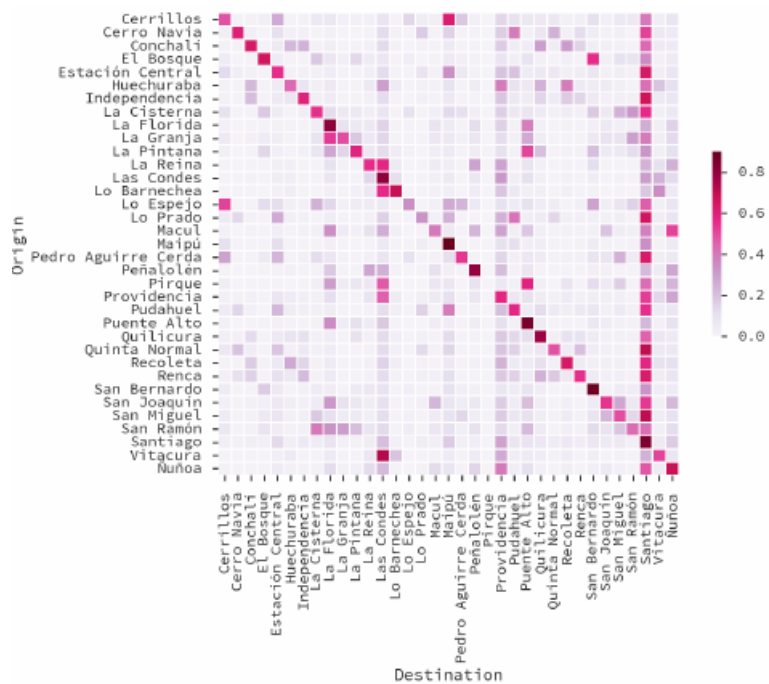
Slika B.2 Primjer mozaika *krhotina* za interpolaciju između Voronoi ćelija i kvadratne mreže. Nijansa odražava postotak površine zgrada u *krhotini* u odnosu na cijelu Voronoi ćeliju kojoj pripada koji čini *težinski faktor* svake polazišne i odredišne *krhotine* te pripadajućih ("roditeljskih") Voronoi ćelija u izračunu toka koji pripada *krhotini*. Sumom svih tokova *krhotina* jedne kvadratne zone izračuna se konačni tok te kvadratne zone. [?]

B.3 Posljedice objedinjenja prostornih zona

Voronoi ćelije također je moguće objединiti u veće zone. Pritom se može u matricu bilježiti i aproksimacija internog toka, odnosno broja putovanja unutar nove ćelije (dijagonala matrice). Takva putovanja nazivaju se internim putovanjima. Studija [?] je pokazala da sa **smanjenjem rezolucije (visokim stupnjem objedinjenja Voronoi ćelija) stupanj korelacije s POM-ma iz ankete (za isti grad) raste** (POM-e dobivene iz CDR i iz anketa u 4 različita grada). Postotak eksternih putovanja koji se gubi u procesu objedinjenja Voronoi ćelija analiziran je u studijama [?] (Vidi sliku ??) i [?] za POM-e iz regije Picardie u Francuskoj. **Objedinjenjem Voronoi ćelija na razinu *Urban Areas* (područja oko gradova) 85% svih putovanja postaje internim putovanjima, a na razini *Urban Cores* (područja oko većih gradova) čak 97% početno zabilježenih putovanja je interno, te ostaje samo 3% eksternih putovanja [?].**



(a) Izvor podataka anketa



(b) Izvor podataka CDR

Slika B.3 Distribucija CDR putovanja na području Santiaga, Čile. POM je normalizirana prema redovima (polazištima) [?]

Dodatak C

OpenStreetMap

OpenStreetMap (OSM) je svjetski raširen projekt koji kreira i pruža slobodne zemljopisne podatke (zemljovide gradova i naselja) temeljen na volonterskom doprinosu zajednice. Projekt pruža detaljne, ažurne digitalne zemljovide kompatibilne s Geoprostorni informacijski sustav (GIS) aplikacijama. Započet je prije 15 godina u Ujedinjenom Kraljevstvu kao odgovor na tehnička ili pravna ograničenja postojećih "slobodnih" baza prostornih podataka kao što je GoogleMaps. Kartu svijeta na visokoj razini kvalitete i točnosti održavaju zajednički doprinos i međusobna kontrola unosa.

Struktura osnovnih elemenata na toj rasterskoj karti je hijerarhijska. Element u hijerarhiji može biti: čvor (engl. *node*), put (engl. *way*) ili relacija (engl. *relation*). Uz hijerarhijske postoji i opisni element koji se naziva oznakom (engl. *tag*), a njegova funkcija je opisati značajke hijerarhijskog elementa uz koji je vezan.

Čvor je jedinstvena točka u prostoru (sa identifikacijskom oznakom, zemljopisnom širinom i dužinom) koja najčešće predstavlja fizički objekt (zgrada, dio ceste...) te sadrži jedan ili više *ključ = vrijednost* oznaka (engl. *key = value tag*) koji definiraju razne značajke objekta.

Put je uređena lista 2 do 2,000 čvorova, koja također može biti opisana *ključ = vrijednost* oznakama.

Relacija je uređena lista čvorova, puteva i/ili relacija. Definira logičku ili geografsku povezanost članova.

Primjeri oznaka vezanih za čvorove: *office=company*, *building=residential*, *building=hotel*, *building=church*, *leisure=sports_centre*, *amenity=school*. Primjer žiga vezanog uz put *highway=residential*.

OSM doprinosi istraživanjima analize kretanja stanovništva i kao izvor prometne infrastrukture, njenih meta podataka (kolnik, mostovi, ograničenja brzine i dr.)[?] i podataka o podnoj površini objekata (engl. *floor area*) [?].

Dodatak D

Procjena polazišno-odredišne matrice B

Podatci u formatu ?? su grupirani prema identifikacijskoj oznaci (Taxi Id) i sortirani prema vremenskom žigu (Time). Početak putovanja definiran je kao prelazak statusa (Occupancy Status) iz 0 u 1, a kraj kao prelazak iz 1 u 0. Prostorna podjela POM je na Voronoi ćelije oko tornjeva baznih stanica jer podatci o TAZ nisu javno dostupni. Generirano je 8 POM, a sva putovanja su dodjeljivana periodu u kojem počinju.

	A	B	C	D	E	F
1	Taxi ID	Time	Latitude	Longitude	Occupancy Status	Speed
2	22223	21:09:38	114.138535	22.609266	1	19
3	22223	11:14:18	114.137871	22.575317	0	0
4	22223	01:18:28	114.137131	22.575983	0	0
5	22223	13:11:42	114.136269	22.545851	1	18
6	22223	02:05:47	114.135948	22.578917	0	29
7	22223	02:05:24	114.135834	22.577433	0	20
8	22223	11:37:32	114.135681	22.566566	1	59
9	22223	01:20:59	114.134331	22.579468	0	0
10	22223	10:47:38	114.132233	22.576866	0	27

Slika D.1 Format podataka iz kojih je procijenjena POM B