

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

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

Rijeka, lipanj 2019.

Vjera Turk
0069064924

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

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

Mentor: izv.prof.dr.sc. Renato Filjar

Rijeka, lipanj 2019.

Vjera Turk
0069064924

Umjesto ove stranice umetnuti zadatak
za završni ili diplomski rad

Izjava o samostalnoj izradi rada

Izjavljujem da sam samostalno izradio ovaj rad.

Rijeka, lipanj 2019.

Ime Prezime

Zahvala

Sadržaj

Popis slika	ix
Popis tablica	x
1 Uvod	1
2 Polazišno-odredišna matrica	2
2.1 Tradicionalni pristupi generiranju POM-a	3
2.1.1 Ankete	3
2.1.2 Prebrojavanje vozila	3
2.1.3 Modeliranje prometa	4
2.1.4 Problematika i ograničenja tradicionalnih načina	4
2.1.5 Pokretna osjetila	5
2.2 Matrice iz zapisa o telekomunikacijskim aktivnostima u javnoj pokret- noj mreži	5
2.2.1 Putovanje	7
2.2.2 Kontekst iz samog izvora podataka o kretanju	7
2.2.3 Moduli kretanja	8
2.3 Kontekst iz vanjskih izvora	8
2.3.1 <i>Points of Interest</i>	9

Sadržaj

2.3.2	OpenStreetMap	9
2.4	Geometrija prostorne podjele	10
2.5	Vremenska domena	16
2.6	Dobre prakse u generiranju POM iz CDR	16
2.7	Drugi primjeri automatskog prikupljanja	17
2.7.1	Dodatni telekomunikacijski podatci ?	17
2.7.2	Združena očitavanja prijamnika za satelitsku navigaciju	17
2.7.3	Javni prijevoz i <i>pametne kartice</i> (<i>Smart Card</i> sustavi)	17
3	Postojeće metrike za validaciju POM-a	18
3.1	Točnost POM-e	18
3.2	Metrike	18
3.3	Strukturalna sličnost	19
3.3.1	MSSI	19
4	Odnosni parametri kvalitete	21
4.1	Zajednički, objektivni kriteriji usporedbe	21
4.2	Komparacijski indikatori	22
4.2.1	Prostorna razlučivost (Rezolucija)	22
4.2.2	Vremenska različitost	22
4.2.3	Širina toka	23
4.2.4	Definicija putovanja	24
4.2.5	Gustoća informacija - kontekst	24
4.3	Međuovisnost parametara	24
	Bibliografija	25
	Pojmovnik	29

Sadržaj

Sažetak	30
A Naslov priloga	31
A.1 Naslov sekcije	31
A.2 Naslov sekcije	31

Popis slika

2.1	Mreža za analizu prometa sastoji se od 221,14 cesta - plave linije i 3,199 Traffic Analysis Zone (TAZ) zona - granice označene iscrtanom crnom linijom [1]	11
2.2	Jedinice lokalne samouprave za grad Shenzhen [2]	12
2.3	Voronoi dijagram baznih stanica - Shenzhen	13
2.4	Podjela na šesterokute Shenzhen - grupacija u različite boje na osnovu širine toka [3]	14
2.5	Prijedlog izmjene (prilagodbe) administrativnih jedinica mreži kretanja - bojom su označene grupacije nastale na temelju matrice kretanja, crnom su označene postojeće granice administrativnih jedinica. [4]	15
4.1	<i>Spaghetti-effect</i> - problem koji se može javiti kod grafičkog prikaza tokova matrice visoke prostorne rezolucije. [5]	23

Popis tablica

Poglavlje 1

Uvod

Izvorišno-Odredišna ili Polazišno-Odredišna matrica *eng. Origin-destination Matrix (ODM)* ili *Trip Table* alat je 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.

Za razliku od tradicionalnog pristupa prebrojavanja putovanja i putnika na licu mjesta, anketiranjem ili pomoću različitih strateški postavljenih osjetila, za procjenu matrica 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 matrice na kontekst. Kada daju informaciju o kontekstu- svrsi kretanja- matrice su dodatno obogaćene i postaju izvor informacija za primjenu i van konteksta prometa.

(...)

Poglavlje 2

Polazišno-odredišna matrica

Polazišno-odredišna matrica sadrži broj putovanja između svakog para zona unutar nekog područja za određeni vremenski okvir (u svakom smjeru zasebno). Redovi u matrici predstavljaju polazišta (izvore prometnog toka), a stupci odredišta. Svaki element u matrici predstavlja broj putovanja iz pripadajuće polazišne u pripadajuću ciljnu zonu. Može se smatrati matričnim zapisom težinskog usmjerenog grafa, gdje su težinski faktori usmjerenih veza širine prometnih tokova između čvorova mreže koji predstavljaju zone na koje je promatrano područje podijeljeno.

U prometnoj znanosti postoje brojne metode estimacije matrica. I dok se još uvijek ulaže u razvoj matematičkih modela koji imaju za cilj povećati kvalitetu matrica generiranih tradicionalnim metodama, u novije vrijeme razvijaju se nove, moderne metode, koje imaju veliki potencijal, zbog svoje jednostavnosti u izvedbi, ažurnosti podataka i velikog uzorka stanovništva koji obuhvate (...)

Matrica može biti generirana za sadašnje ili planirano prometno opterećenje. Može biti prikazana ukupno ili po pojedinim modulima prometnog sustava (osobna vozila, vozila javnog prijevoza, teretna vozila, pješaka itd) i/ili različitim svrhama putovanja (putovanja na posao, putovanja kući...).

Određivanje konteksta putovanja dodatno obogaćuje matrice jer specificira socio-ekonomske aktivnosti koje su pokretači putovanja odnosno razlozi kretanja stanovništva. Mogućnost preciznog određivanja konteksta zavisi o metodi estimacije matrica.

Kada kontekst (svrha) putovanja nije inicijalno poznat iz prikupljenih podataka o putovanjima, razvijene su metode koje pomoću vanjskih izvora o sadržaju procjenjuju kontekst putovanja.

2.1 Tradicionalni pristupi generiranju POM-a

2.1.1 Ankete

Tradicionalne metode uključuju metode provođenja anketiranja. Postoji nekoliko vrsta anketa: u kućanstvima, presretanje vozila, anketiranje prijevoznih kompanija i tvrtki, tranzit na granici, turisti u hotelima i parking [6]. Ovakve metode su prije svega resursno zahtjevne stoga se ne provode često. Anketiranje putovanja u kućanstvima provodi se na područjima brojnih svjetskih metropola tek jednom u deset godina.

Anketiranje vozača presretanjem vozila eng. Road Side Interview (RSI) u prosjeku košta čak 10 eura po jednom zabilježenom putovanju i ukupno obuhvati 10% prometa. Ova vrsta najčešće se provodi na autocestama. [7]

Provođenje ovih metoda zahtjeva mnogo vremena i ponekad su netočne. Pokrivaju malen dio populacije te iz tog razloga mogu biti pristrane.[8]

2.1.2 Prebrojavanje vozila

Metode koje uključuju prebrojavanje vozila na ključnim čvorištima prometne infrastrukture manje su zahtjevne u odnosu na ankete i značajno smanjuju vrijeme i troškove prikupljanja podataka. Radi se o analizi nadzornih snimaka prometa, automatskom sustav za prepoznavanje i očitovanje registarskih pločica eng. Automatic Number Plate Recognition (ANPR), osjetilima prometnog toka (radarski, magnetni, video-senzori, zvučni...) i ručnom prebrojavanju.

2.1.3 Modeliranje prometa

Matematičko 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 POM-u koja specificira prometnu potražnju između zona u prometnoj mreži. Točna Polazišno-Odredišna Matrica (POM)-a osnova je za odluke u mnogim Inteligentnim Transportnim Sustavima engl. Intelligent Transportation Systems (ITS) koji su zauzvrat ključni za izbor ruta u različitim sustavima navigacijskog navođenja. Neke od pretpostavki kojima se vode u modeliranju su primjerice da je ukupna dnevna potražnja podjednaka u oba smjera kod svakog para zona na području grada. [9]

2.1.4 Problematika i ograničenja tradicionalnih načina

Širenje gradova i rast stanovništva rezultirali su rastućim brojem sve ozbiljnijih prometnih zastoja u velikim gradovima diljem svijeta. Prepoznata je potreba za opsežnim strategijama upravljanja prometom kako bi se suočili s izazovima koje donosi brzo razvijajuća okolina i demografija populacije. Efektivno upravljanje i kontrola prometa mogu povećati sigurnost, kvalitetu usluge, poticati ekonomski rast i smanjiti zagađenje zraka. Zbog dinamike kojom se gradovi mijenjaju razumno je pitati se ima li smisla koristiti matrice stare više godina za modeliranje današnjeg prometa. Dobivanje što točnijih i jeftinijih POM-a privlači pažnju znanstvenika i van područja prometne znanosti. Alternativni izvori koje oni predlažu za promatranje obrazaca putovanja su značajno jeftiniji, ali nailaze na problem da profesija zahtijeva da se njihova vrijednost dokazuje usporedbom s tradicionalnim metodama. [7]

2.1.5 Pokretna osjetila

Osnovna ideja pokretnih osjetila je da vozila koja se kreću po prometnoj mreži i sastavni su dio prometnog toka prikupljaju i šalju podatke. Nužan zahtjev koji treba biti ispunjen kod pokretnih osjetila je poznavanje pozicije vozila u određenim vremenskim trenucima. Za to se koristi Global Navigation Satellite System (GNSS) ili se vozila identificiraju na određenim točkama. Engleski naziv za ovu metodu je Floating Car Data (FCD). Danas je FCD komplementarni izvor vrijednih podataka i postaje najvažnija tehnologija prikupljanja podataka inteligentnih transportnih sustava. U prikupljanju podataka sudjeluju “namjenska” vozila (taksi, distribucija roba, javni prijevoz, službena vozila i dr.), a mora biti uspostavljen način učestalog slanja podataka o položaju a najčešće se odvija preko Global System for Mobile (Communications) (GSM), Wi-fi pa i Bluetooth tehnologije. [10]

Ukoliko pokretno osjetilo nije vozilo već pametni telefon tada se često koristi naziv Floating Phone Data (Floating Cellular Data) (FPD). FPD se odnosi na prikupljanje podataka pomoću mobilnih (pametnih) telefona neovisno o načinu određivanja položaja - korištenjem GNSS ili pomoću javne pokretne mreže.

Zaseban pristup unutar FPD je određivanje položaja unutar javne pokretne mreže (mreže baznih stanica). Kod ovog načina položaj pokretne stanice može se odrediti triangulacijom (pomoću signala iz više baznih stanica čije signal hvata pokretna stanica) ili se aproksimira područjem pokrivanja bazne stanice na koju je u trenutku očitavanja položaja pokretna stanica spojena.

2.2 Matrice iz zapisa o telekomunikacijskim aktivnostima u javnoj pokretnoj mreži

Kako bi pružali i naplaćivali usluge, mobilni operateri moraju prikupljati podatke o pokretnim stanicama korisnika. Pojednostavljeno, da bi se usmjerio poziv do odgovarajućeg uređaja, mora se biti poznato s koje bazne stanice odaslati signal. Naziv za podatke koje operateri prikupljaju kako bi naplaćivali usluge glasi Call Data Records ili Charging Data Records (CDR). Radi se o zapisima aktivnosti u mobilnoj mreži

Poglavlje 2. Polazišno-odredišna matrica

- pozivima, porukama (engl. Short Messaging Service - SMS) te započetom prijenosu mobilnih podataka. Tipičan zapis između ostalog sastoji se od identifikacijske oznake pristupne točke radijske mreže (bazne stanice), vremenskog žiga početka telekomunikacijske aktivnosti i trajanja aktivnosti. Istraživanje veza između kretanja i društvenih veza (poznanstava) [11] od operatera je dobilo podatke i na koju baznu stanicu je poziv usmjeren.

Dodatni podatci koje operateri prate (*Signaling data*) prikupljani su pomoću sustava za ispitivanje kvalitete mreže. Uključuju *handover* podatke - podatke o transferu na drugu baznu stanicu tijekom trajanja aktivnosti (poziva), *Location Area updates* - podatke prokupljane svakih 20 min do 2 sata o tome na koju skupinu baznih stanica je pokretna stanica spojena, te informacije o tome kada je pokretna stanica uključena u mrežu i isključena iz mreže. Pri izradi matrica, ako su dostupni, svi ovi podatci mogu se uzeti u obzir i poboljšati model.

U novije vrijeme znanstvenici predlažu metode u kojima koriste upravo anonimizirane CDR kao izvor podataka za generiranje POM-a. [12] [8] [13] iako se sama ideja spominje u istraživanjima još iz 1994., gotovo od kad je javna pokretna mreža postala široko dostupna. [13] Ako u daljnjem tekstu nije spomenut izvor, radi se o matricama generiranim iz CDR.

Zanimljivo je ovdje spomenuti europsku inicijativu *eCall* [14] za brzu pomoć stradalima u prometnim nesrećama bilo gdje u Europskoj Uniji. *eCall* zahtjeva da u svakom novom motornom vozilu bude ugrađen uređaj koji u slučaju prometne nesreće automatski obavijesti centar 112 te mu šalje informacije o aktivaciji zračnih jastuka, prikupljene podatke osjetila za procjenu jačine sudara i u dakako položaj vozila u trenutku prometne nesreće. Također treba omogućiti uspostavu poziva (GSM). Prognoza je da će se korištenjem tog sustava vrijeme odziva (dolaska službi na mjesto prometne nesreće) smanjiti za između 40% i 50%. Od travnja 2018. svi novi automobili u prodaji u Europskoj Uniji imaju ugrađen ovaj uređaj. Nakon uspješnog uspostavljanja sustava očekuje se da bi se uređaj mogao početi koristiti i u druge svrhe.

Danas uz sveprisutnost i prodornost tehnologije, podatci koje generiraju pametni telefoni omogućuju podatkovnim analitičarima razumjeti ponašanje u brojnim domenama pojedinačno, uključujući i njihov obrazac kretanja. Kao što je već spomenuto

Poglavlje 2. Polazišno-odredišna matrica

matrice je moguće generirati iz CDR koje operateri prikupljaju u svrhu pružanja i naplaćivanja usluge. Za ovu metodu stoga ne postoji potreba za dodatnim uređajima, sustavima ili protokolima centralnog prikupljanja položaja kao što je to u slučaju kada se položaj određuje pomoću GNSS prijamnika.

Pokriva pješački promet, (...anonimizacija i k-anonimizacija...).

2.2.1 Putovanje

Priroda ovog pristupa rudarenju podataka za estimaciju POM-a zahtjeva da prvo definiramo što točno smatramo putovanjem, odnosno polazištem i odredištem putovanja. Zaustavljanje se definira kao skup minimalno n događaja (aktivnosti) koji su zabilježeni na istoj poziciji unutar definiranog minimalnog vremenskog perioda. Time se osigurava da se ne radi o tranzitnoj već o namjereoju poziciji zaustavljanja (i suzbija se bilježenje lažnog kretanje kod privremenog spajanja na susjednu bazni stanicu). Algoritam u nizu aktivnosti svakog pojedinca traži takva zaustavljanja. Dobiveni niz kontinuiranog zaustavljanja na različitim položajima otkriva kretanje. Algoritam izvlači uzastopne parove zaustavljanja te, prema dodatno definiranim kriterijima dozvoljenog minimalnog i maksimalnog vremenskog razmaka između polaska i dolaska na odredište, ranija ćelija zaustavljanja postaje polazište, a kasnija odredište putovanja. Skupovi putovanja u pripadajućem vremenskom okviru sačinjavaju matricu za taj vremenski okvir.

2.2.2 Kontekst iz samog izvora podataka o kretanju

Kada su dostupni zapisi prikupljeni dulje vremensko razdoblje, tjednima ili mjesecima često se putovanja odnosno matrice mogu definirati na drugačiji način. Pojednostavljeno, analizom zapisa može se utvrdi na kojem položaju (u kojoj ćeliji) neki pojedinci pretežito borave u tipičnom radnom vremenu te u kojoj tokom noći. Prva ćelija za njihova putovanja postaje *Work* destinacija, druga *Home*, a sve preostale destinacije svrstaju se u zasebnu kategoriju *Other*. Potom se generiraju matrice koje uključuju samo putovanja iz jednog tipa destinacije u drugi npr. Home-Work (HW) i Work-Home (WH). [15] [16][17] [18] [5]. Jedan od razloga je da su takve matrice

Poglavlje 2. Polazišno-odredišna matrica

usporedive s onima dobivenim iz anketa stanovništva s istom podjelom putovanja.

Bitno je naglasiti da kod ovih vrsta matrica najčešće naglasak nije stavljen na dobru vremensku rezoluciju.^(*)[13]

- Duža razdoblja prikupljanja podataka omogućuju uočavanje uzoraka kretanja i motiva (teorija mreža). [19]

- [5] str. 4 podjela trip-based, activity-based i another approach

2.2.3 Moduli kretanja

Isključivo CDR zapisi ne osiguravaju rezoluciju pogodnu za određivanje modula kretanja iako su razvijene metodologije [18]. Iznimno se mogu izdvojiti bazen stanice dedicerane pružanju usluge isključivo u podzemnoj željeznici.

Terralytics je jedna od mladih kompanija koja nudi uslugu modeliranja prometa iz telekomunikacijskih podataka. Između ostalog prema brzini kojom se pokretna stanica (ili skupina njih istovremeno) mijenja položaj donose zaključak o dominantnom modulu kretanja (pješač, automobil, javni prijevoz, avion). Tajna njihovog uspjeha je da osim CDR koriste i ranije spomenute dodatne podatke koji se prikupljaju. [20]

Za razliku od tradicionalnih modela prebrojavanja vozila, telekomunikacijski zapisi obuhvaćaju sve module kretanja kao i svu prometnu infrastrukturu, ne samo onu gdje se provode intervjui ili gdje su postavljena osjetila za prebrojavanje prometa. Uklanja potrebu za različiti pristup prebrojavanja za pojedini modul kretanja.

2.3 Kontekst iz vanjskih izvora

Neka istraživanja spojila su ljudsku mobilnost s namjenom zemljišta (land use) ili točkama interesa *engl. Points of Interest (POIs)* kako bi grupirala urbane predjele prema njihovoj funkciji ili upotrebi. Izvori podataka o mobilnosti variraju od Global Positioning System (GPS) putanja, položaja ukrcaja i iskrcaja taxi vozila, zapisi koje generiraju pametne kartice u javnom prijevozu, podatci o dijeljenim lokacijama na društvenim mrežama (*engl. check-in*) i zapisi o telekomunikacijskim aktivnostima (CDR). [8]

2.3.1 *Points of Interest*

U jednom od svojih radova grupa autora kvantizira vezu između ukupno 23 namjenske kategorije POIs i onim što nazivaju 3 tipa privlačenja (...nabrojiti neke kategorije). Svaki tip privlačenja ima karakterističan tok (gustoću ukupnog toka, udaljenost i disperziju polazišta tokova usmjerenih prema objektu tog tipa privlačenja). [8] U radu koriste podatke iz službene baze s 12,000 POIs grada u pitanju i CDR zapisima iz razdoblja od mjesec dana.

2.3.2 OpenStreetMap

OpenStreetMap (OSM) je svjetski raširen projekt koji kreira i pruža slobodne zemljopisne podatke (zemljovide gradova i naselja) temeljen na volonterskom doprinosu zajednice. Pruža detaljne i ažurne digitalne zemljovide kompatibilne s Geographic Information System (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 Google Maps. Zajednički doprinos i međusobna kontrola unosa održavaju kartu na visokoj razini kvalitete i točnosti.

Izvor infrastrukture

Jedan od načina na koji OSM doprinosi istraživanjima u ovom području je kao izvor prometne infrastrukture i pripadajućih meta podataka (kolnik, mostovi, ograničenja brzine...). Raspodjela toka po prometnoj infrastrukturi nadilazi područje ovog rada no jest dio problematike modeliranja prometa. [18]

Model raspodjele toka

Inovativni pristup kontekstualizaciji toka korištenjem OpenStreetMap baze predstavljen je u radu [21]. Na osnovu pripadajućih opisnih podataka, prostorni objekti na području interesa kategorizirani su prema tipu socio-ekonomske aktivnosti u koju su uključeni. Definirano je ukupno 6 kategorija: Dom (Home), Posao (Work), Zdravlje (Health), Edukacija (Education), Zabava (Leisure) i Ostalo (Other). Razvijen je

vjerojatnosni model koji, na osnovu broja objekata pojedine kategorije u odredišnoj ćeliji i promatranog vremenskog okvira, ukupni ulazni tok u odredišnu ćeliju dijeli na 6 tokova usmjerenih prema objektima tih kategorija. Iz jedne matrice tako se dobije 6 matrica, po jedna za svaku od kategorija. Korištene su matrice dobivene iz javno dostupnih, anonimiziranih telekomunikacijskih zapisa na području kineskog grada Shenzhena. [22][23] Na ovaj način definicijom putovanja sačuvana je mogućnost dobre rezolucije u vremenu a istovremeno je dan kontekst koju uključuje *home* i *work* kategorije.

2.4 Geometrija prostorne podjele

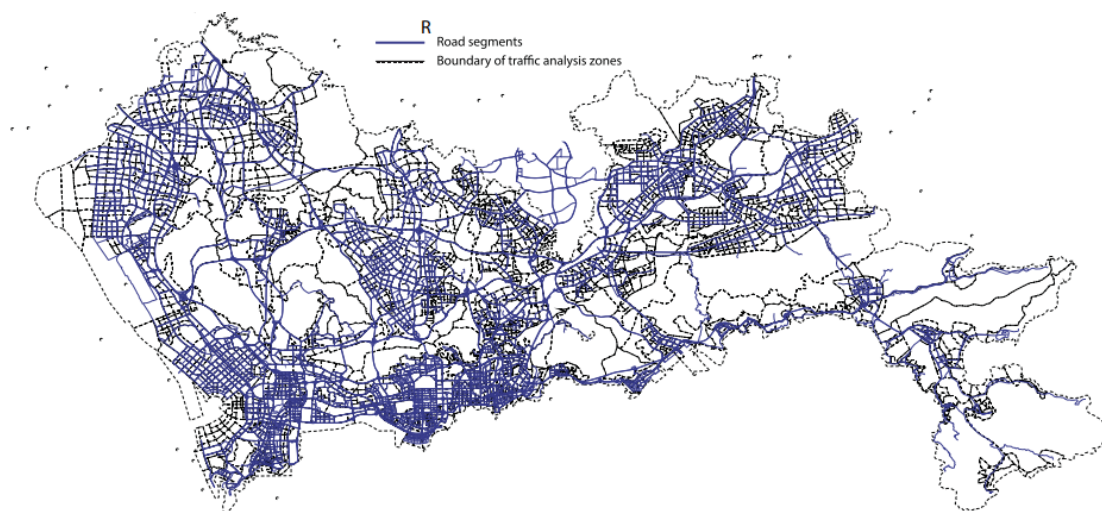
Konačna geometrija prostorne podjele alternativnih izvora uglavnom ovisi o prostornoj podjeli *grand truth* matrice. Obično se radi o jedinici prostorne podjele naziva Traffic Analysis Zone (TAZ) kod analize prometa, o anketnim jedinicama ili jedinicama samouprave. Neprihvaćenija prostorna podjela kod matrica izvedenih iz CDR je ona na Voronoi ćelije.

Traffic Analysis Zone (TAZ)

TAZ je strani naziv za jedinicu prostorne podjele (ćeliju) kod analize prometa. Kod programskih podrška za analizu prometa veličina jedinica varira i nije uniformna. Podjela se može temeljiti na populaciji ili broju zaposlenih na nekom prostoru, primjerice svaka ćelija pokriva prostor na kojem obitava otprilike tri tisuće ljudi (propisanih preporučenih vrijednosti nema), ili podjela u kojoj ćelije generiraju podjednak broj izlaznih putovanja (jednaka širina izlaznih tokova) [15]*. U hijerarhijskoj podjeli kod ankete stanovništva općenito postoje manje i veće jedinice podjele, a nema podjele koja bi precizno odgovarala TAZ. Do 2000. godine zastupljen je bio jedan model predviđanja prometa, a veličina TAZ bila je uvijek približno jednaka anketnom bloku od 600 do 3000 ljudi. Pojavom kompleksnijih modela predviđanja prometa (i modela upotrebe zemljišta) počinju se koristiti sve manje prostorne jedinice jer se povećava i razina detalja koju noviji modeli mogu ostvariti. Jedinica veća od TAZ uvedena 2010. naziva se Traffic Analysis District (TAD) koji obuhvaća oko 20,000

stanovnika.[24]

Kod analize na području cijele države [15] koristi *k-sredina* algoritam za grupiranje baznih stanica u broj grupa jednak broju TAZ na području Senegala.



Slika 2.1 Mreža za analizu prometa sastoji se od 221,14 cesta - plave linije i 3,199 TAZ zona - granice označene iscrtanom crnom linijom [1]

Prema podacima korištenim u istraživanju iz 2018. [25] Shenzhen je podijeljen na 996 TAZ od kojih 844 zone imaju barem jednu autobusnu ili stanicu podzemne željeznice. Godinu dana ranije istraživanje spominje više zona (Vidi sliku 2.1).

Jedinice samouprave i anketne jedinice

[26] ne definira unaprijed broj grupa već za grupaciju koristi *mean-shift* algoritam koji sam određuje broj grupa na području Švedske, potom koristeći OSM grupe veže uz imena gradova i mjesta. U istraživanju [11] bazne stanice na području Kolumbije grupiraju po jedinicama samouprave (eng. *municipalities*). Isti postupak je korišten i u istraživanju na području Madrida, kako bi matrice u konačnici bile usporedive s onima dobivenim iz anketa. [12]. Prikaz jedinica samouprave za grad Shenzhen možemo vidjeti an slici 2.2

Poglavlje 2. Polazišno-odredišna matrica



Slika 2.2 Jedinice lokalne samouprave za grad Shenzhen [2]

Voronoi teselacija

Najfinija rezolucija koju je moguće postići kada je riječ o matricama dobivenim iz CDR je Voronoieva teselacija.

Općenito, u ravnini, za svaku točku iz skupa točaka postoji odgovarajuća regija koju čine sve točke ravnine bliže toj točki negoli ikojoj drugoj iz skupa. Te regije se nazivaju Voronoievim ćelijama. 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.

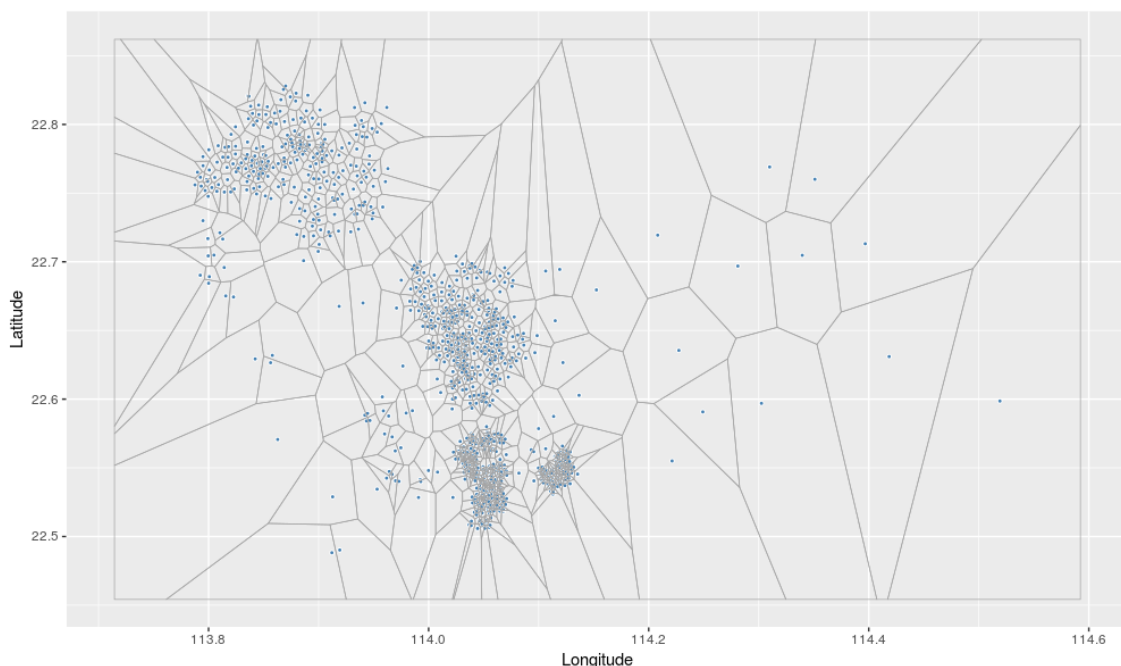
Točke skupa su položaji baznih stanice, a ćelije oko njih aproksimacija površine koju ta bazna stanica pokriva. U stvarnosti se te površine prekrivanja preklapaju. (Vidi sliku 2.3)

Ponekad operateri u CDR zapisima ne otkrivaju položaje baznih stanica već samo njihov ID. Postoji više javno dostupnih baza nastalih volonterskim doprinosom koje pohranjuju podatke o procjenama položaja baznih stanica, snazi signala i sl. Položaji

Poglavlje 2. Polazišno-odredišna matrica

se prikupljaju pomoću aplikacije instalirane na mobilnom uređaju, registriranjem GPS pozicije i identifikacijske oznake bazne stanice na koju je uređaj trenutno spojen. Dakle, ove baze ne sadrže "najtočniji" položaje baznih stanica već usrednjene položaje mobilnih uređaja koji su u nekom trenutku bili spojeni na tu baznu stanicu. Može se reći da taj procijenjeni položaj za potrebe izrada matrica čak i više odgovara od stvarnog položaja. [26]

Gustoća Voronoi ćelija snažno je korelirana s gustoćom stanovništva te stoga bolju prostornu rezoluciju ima na područjima veće gustoće populacije. No Gundlegard [13] smatra da neuniformnost ćelija nije problem, već preklapanje područja pokrivanja. Posljedica toga su stalne oscilacije - uređaj mijenja baznu stanicu na koju je spojen iako svo vrijeme zadržava isti položaj. Jedan od načina na koji se može ublažiti efekt koji oscilacija ima na definiranje putovanja je odrediti minimalnu udaljenost putovanja ili potpuno zanemariti putovanja u susjedne ćelije.



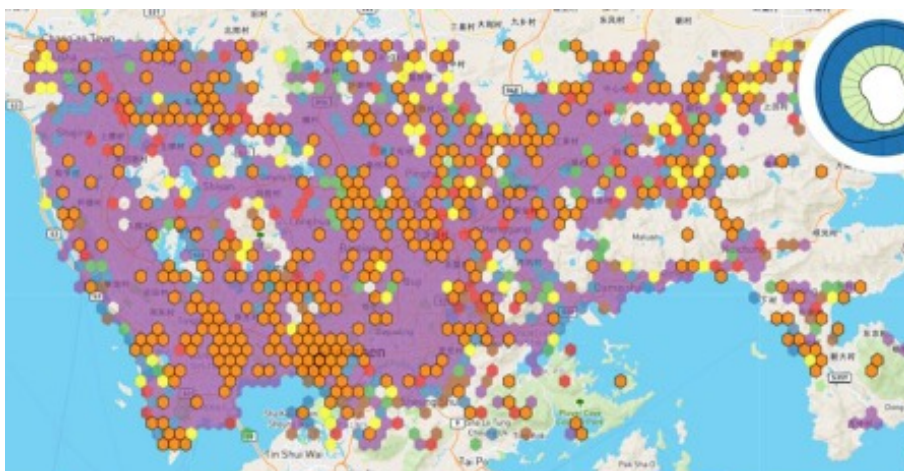
Slika 2.3 Voronoi dijagram baznih stanica - Shenzhen

Pravokutna mreža i mreža šesterokuta

Uniformna podjela na pravokutne blokove od 1 kvadratni kilometar spominje se kao dobra praksa [15] (...)

Područje pokrivanja bazne stanice na grafikama se standardno prikazuje kao šesterokut (heksagon). [20] između ostalih podjela koristi i šesterokute koje po potrebi ugnježđuje i tako mijenja rezoluciju u kojoj promatra gibanja. Tvrdi da su se u nekim segmentima pokazali i 13% bolji od pravokutne mreže.

Prikaz šesterokuta za grad Shenzhen možemo vidjeti na slici 2.4



Slika 2.4 Podjela na šesterokute Shenzhen - grupacija u različite boje na osnovu širine toka [3]

Interpolacija i usporedba

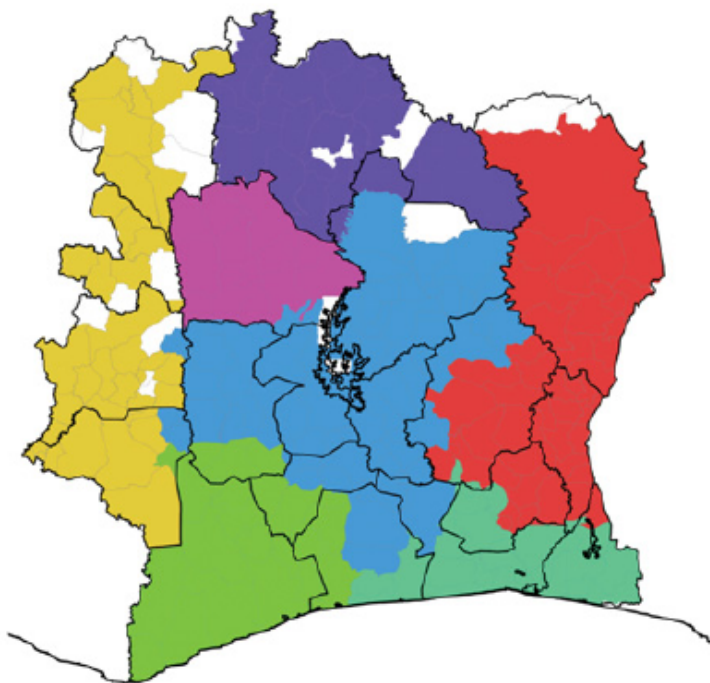
[18] [5]

Redefiniranje prostorne podjele na temelju matrica

U sklopu projekta Data for Development (D4D) francuski operater *Orange* ustupio je na korištenje CDR višemjesečne zapise s područja Obale Bjelokosti. Različita istraživanja u ovom području rađena su upravo na tim podacima. Jedno od njih [11]

Poglavlje 2. Polazišno-odredišna matrica

ističe da je dobar broj istraživanja dokazao da administrativne regije u zemlji faktor koji definira ljudsku mobilnost i interakcije na području države. Da bi testirali tu tvrdnju oni, nakon što generiraju matricu kretanja, ćelije prostorne podjele grupiraju prema *community detection* algoritmu. Iako se dio grupa vrlo lijepo preklapa sa postojećim administrativnim jedinicama, one administrativne jedinice u kojima se nalaze veliki gradovi, grupacije nastale na temelju mobilnosti bitno su drugačije. Zaključak je da potencijalno postoji potreba da se granice administrativnih jedinica izmjene (Slika 2.5)



Slika 2.5 Prijedlog izmjene (prilagodbe) administrativnih jedinica mreži kretanja - bojom su označene grupacije nastale na temelju matrice kretanja, crnom su označene postojeće granice administrativnih jedinica. [4]

2.5 Vremenska domena

-tranzitna matrica?,matrica bez vremenske dimenzije (odnosno ima samo jedan vremenski okvir) -radni dan, vikend, petak drugačiji -i, sposobnost CDR da uhvate razlike -sposobnost da uhvati sezonske promjene

2.6 Dobre prakse u generiranju POM iz CDR

-k-anonymization

-Grad, Država

A potential is also recognized for a planet-scale mobility measurement through opening to researchers and combining the different big dataset carrying location information.[4]

-CDR- POM u zemljama u razvoju

-Penetracija od 82%, 9x% -i, pronaći radove

-Mobilni uređaji iz godine u godinu zastupljeni su kod sve većeg postotka stanovništvu, ne samo u razvijenim zemljama već i u zemljama u razvoju.

2.7 Drugi primjeri automatskog prikupljanja

2.7.1 Dodatni telekomunikacijski podatci ?

2.7.2 Združena očitavanja prijamnika za satelitsku navigaciju

komercijalna vozila, Taxi

Jedno istraživanje matrica iz cdr je eliminiralo zapise onih korisnika koju su radili dulja "neprekidna putovanja" pod napomenom da se radi o taksistima i sl. te da ih neće uzeti u obzir.

2.7.3 Javni prijevoz i *pametne kartice* (*Smart Card* sustavi)

82% putovanja javnim prijevozom naprave korisnici javnog prijevoza sa pametnim karticama. [27]

Poglavlje 3

Postojeće metrike za validaciju POM-a

3.1 Točnost POM-e

Točnost procijenjenih matrica gotovo uvijek se definira u odnosu na referentnu matricu (eng. *grand truth matrix*) koja je dobivena tradicionalnim postupcima (anketiranje i/ili prebrojavanje vozila). Statističke mjere kvantiziraju razliku procijenjenih i “istinitih” vrijednosti, ako su nam one poznate.

Često se u literaturi (jednoznačno) koriste pojmovi *točnost*, *pouzdanost* i *kvaliteta*. Gotovo uvijek radi se o mjerama koje opisuju razinu sličnosti odnosno razlike (greška) s referentnom matricom .

3.2 Metrike

Za procjenu kvalitete matrica dobivenih isključivo anketranjem u radu [6] korištena je mjera Mean Absolute Percentage Error (MAPE), te je prikazano da se zadovoljavajuća razina kvalitete takvih matrica postiže tek ako uzorak obuhvaća 50% populacije. Istaknuta je važnost korištenja dodatnih izvora za izradu matrica.

U radu [28] navedene su statističke mjere Relative Error (RE), Total Demand

Deviation (TDD), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) te Maximum Possible Relative Error (MPRE) i 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). [29]

U [12] korišten je *Pearsonov koeficijent korelacije* - r da bi se utvrdila sličnost svakog retka matrice dobivene iz CDR s retkom referentne (izlazni tok iz svake polazišne ćelije). Isti postupak korišten je za kontekstualizirane HW i WH matrice dobivene iz CDR u usporedbi s referentnim matricama dobivenim anketiranjem.

Travassoli u svome radu [27] navodi nekoliko uobičajeno korištenih mjera - R^2 , Geoffrey E. Havers statistics (GEH), Root Mean Squared Error percentage %RMSE te uvodi novu mjeru Eigenvalue-based measure (EBM) (temeljenu na svojstvenim vrijednostima matrica) i procjenjuje pouzdanost matrice dobivene iz sustava automatskog prikupljanja podataka u javnom prijevozu (autobus, vlak i trajekt). Spominje i Wasserstein metric, mjeru koja se razlikuje po tome da ne uspoređuje samo vrijednosti parova istih ćelija (elementwise).

Spearmanov koeficijent korelacije ranga korišten je u [17] za procjenu sličnosti matrica dobivenih iz CDR sa tada aktualnim matricama dobivenim anketiranjem.

3.3 Strukturalna sličnost

Dosada spomenute mjere neće uhvatiti strukturalnu sličnost matrica. Nekolicina autora ističe važnost strukturalne sličnosti s referentnom matricom kao važnu mjeru kvalitete matrice jer visoka razina strukturalne sličnosti može biti prisutna i kod matrica s manjom razinom sličnosti prema statističkim mjerama. Također, strukturalna sličnost je (vizualno) vidljiva u grafičkom obliku matrice.

3.3.1 MSSI

Mean Structural Similarity index (MSSIM) dolazi iz područja računalne obrade slike i koristi se kao mjera usporedbe digitalnih slika (*eng. measure of comparison*). U prometu ideja o korištenju MSSIM za mjerenje sličnosti matrica se prvi puta spominje

Poglavlje 3. Postojeće metrike za validaciju POM-a

i demonstrira na simuliranim matricama dobivenim iz referentne matrice dodavanjem šuma. [29]

Structural Similarity index (SSIM) bazira se na degradaciji strukturalnih informacija na jednoj slici u usporedbi s drugom (referentnom) slikom. SSIM se računa za svaki blok veličine $N \times N$ na način da se jezgra (da bi obuhvatila novi blok) pomiče ćeliju po ćeliju dok ne prođe preko cijele slike. MSSIM je srednja vrijednost svih SSIM.

osnovni

[29]

poboljšani

[7][30]

Poglavlje 4

Odnosni parametri kvalitete

Parametar - varijabla o kojoj ovisi određeni logički izraz, matematička formula ili funkcija, a koju promatramo kao dodatnu ovisnost u izrazu koji se definira kao da je ta vrijednost čvrsta.

4.1 Zajednički, objektivni kriteriji usporedbe

- Isti grad
- Isto doba godine
- Isto vremensko razdoblje -Week day, work day

Kod generiranja matrica treba definirati hoće li se putovanja koja se protežu kroz više perioda dodijeliti vremenskom periodu u kojem započinju ili u kojem završavaju. Iako ponekad nije specificirano, [28] [22] sva putovanja dodjeljuju intervalu u kojem je putovanje započeto. citeNetko:nekad Ne zanemaruje one zapise iz kojih nije jasno vidljiv početak putovanja već takva putovanja dodjeljuje na osnovu vjerojatnosti da je započelo u određenom periodu.

-Ima li matrica diagonalu? - kada se preslikava bs ćelije na drugi oblik, nastaju putovanja unutar novih ćelija (koje su veće od bs ćelija) [17]

4.2 Komparacijski indikatori

4.2.1 Prostorna razlučivost (Rezolucija)

Prema hrvatskoj enciklopediji definicija razlučivosti (rezolucije) glasi: mjera za razaznavanje sitnih pojedinosti na nekom prikazu (npr. televizijskoj slici). U računalstvu se odnosi na finoću rasterske slike iskazanu ukupnim brojem slikovnih elemenata (relativna razlučivost) ili brojem slikovnih elemenata po inču (stvarna razlučivost). (...) Kod POM ...

Rezolucija će ovisiti o točnosti položaja ...

Voronoi ćelije također je moguće ugnijezditi (što se često radi kod usporedbe s matricama dobivenim drugim postupkom), a pritom se može u matricu zabilježiti i aproksimacija internog toka, odnosno broja putovanja unutar nove ćelije (popuniti dijagonalu). Postotak eksternih putovanja koji se gubi u procesu agregacije Voronoi ćelija analiziran je za matrice iz regije Picardie u Francuskoj [5].

Agregacijom na razinu *Urban Areas* (gradova) 85% svih putovanja postaje internim putovanjima, a na razini *Urban Cores* čak 97% početno zabilježenih putovanja je interno, te ostaje samo 3% eksternih putovanja.

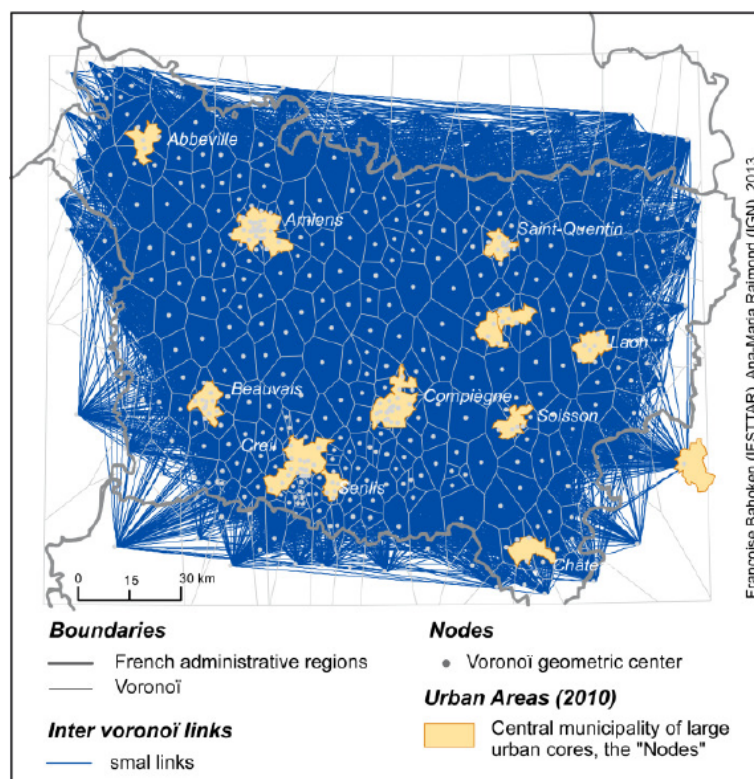
Ruralna područja rjeđe zahtijevaju agregiranje ćelija.

Na temelju usporedbi matrica kretanja u 4 različita grada, dobivenih iz CDR i iz anketa, gdje je svaki grad u anketnim matricama imao svoju prostornu podjelu odnosno rezoluciju, Toole u svom radu [18] (očekivano) zaključuje da sa **smanjenjem rezolucije (visokim stupnjem agregacije Voronoi ćelija) stupanj korelacije s matricama iz ankete raste.**

4.2.2 Vremenska različitost

Departure/Arrival time

Poglavlje 4. Odnosni parametri kvalitete



Slika 4.1 *Spaghetti-effect* - problem koji se može javiti kod grafičkog prikaza tokova matrice visoke prostorne rezolucije. [5]

Dani u tjednu - radni dani/petak/vikend

Sezonske razlike

4.2.3 Širina toka

Ukupan broj odlazaka/dolazaka po vremenskom okviru za cijelu matricu.

-Obuhvaća pješački promet

- veličina uzorka (postotak stanovništva)

4.2.4 Definicija putovanja

prosječan broj putovanja po stanovniku/ispitaniku

Infrastruktura

Obuhvaća pješački promet

4.2.5 Gustoća informacija - kontekst

4.3 Međuovisnost parametara

Ukoliko je rezolucija mala (velike ćelije) nema potrebe za preciznim definiranjem kraja

Bibliografija

- [1] J. Xie, Y. Nie, and X. Liu, “Testing the proportionality condition with taxi trajectory data,” *Transportation Research Part B Methodological*, May 2017.
- [2] M. K. Ng and W. S. Tang, “The Role of Planning in the Development of Shenzhen, China: Rhetoric and Realities,” *Eurasian Geography and Economics*, July 2014.
- [3] Z. Zhou, J. Yu, Z. Guo, and Y. Liu, “Visual exploration of urban functions via spatio-temporal taxi OD data,” *Journal of Visual Languages & Computing*, vol. 48, pp. 169–177, October 2018.
- [4] S. Scepanovic, I. Mishovski, P. Hui, J. K. Nurminen, and A. Yla-Jaaski, “Mobile Phone Call Data as Regional Socio-Economic Proxy Indicator,” *PLoS ONE*, 2015.
- [5] F. Bahoken and A.-M. Olteanu-Raimond, “Designing Origin-Destination Flow Matrices from Individual Mobile Phone Paths: The effect of spatiotemporal filtering on flow measurement,” in *ICC 13 - 26th International Cartographic Conference, Aug 2013, DRESDEN, Germany. ICC 13 - 26th International Cartographic Conference 15p*, 2013., hal-01011987v2.
- [6] M. Cools, E. Moons, and G. Wets, “Assessing the Quality of Origin-Destination Matrices Derived from Activity Travel Surveys,” *Transportation Research Record: Journal of the Transportation Research Board*, 2010.
- [7] T. Pollard, N. Taylor, T. van Vuren, and M. MacDonald, “Comparing the Quality of OD Matrices in Time and Between Data Sources,” *European Transport Conference*, 2013.
- [8] M. Alhazzani, F. Alhasoun, Z. Alawwad, and M. C. González, “Urban Attractors: Discovering Patterns in Regions of Attraction in Cities,” *Public Library of Science*, 2016.

Bibliografija

- [9] A. Peterson, “The Origin-Destination Matrix Estimation Problem- Analysis and Computations,” Ph.D. dissertation, Linköping Studies in Science and Technology, 2007.
- [10] N. Jelusic, “Telematicka sucelja (nastavni tekst),” 2016.
- [11] M. Coscia and R. Hausmann, “Evidence that Calls-based Mobility Networks are Isomorphic,” *Working Papers - Cernter for International Development at Harvard University*, 2015.
- [12] V. Frías-Martínez, E. Frías-Martínez, and C. S. Ruiz, “Estimation of urban commuting patterns using cellphone network data,” 2012.
- [13] D. Gundlegard, C. Rydergren, N. Breyer, and B. Rajna, “Travel demand estimation and network assignment based on cellular network data,” *COMPUTER COMMUNICATIONS*, 2016.
- [14] (06.05.2019) ecall. , s Interneta, <https://en.wikipedia.org/wiki/ECall>
- [15] J. Goulding, *Best Practices and Methodology for OD Matrix Creation from CDR data*, N/LAB, University of Nottingham, 2016.
- [16] F. Calabrese, G. D. Lorenzo, L. Liu, and C. Ratti, “Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area,” *IEEE Pervasive Computing* 10, no. 4, pp. 36–44., April, 2011.
- [17] E. Graells-Garrido and D. Saez-Trumper, “A Day of Your Days: Estimating Individual Daily Journeys Using Mobile Data to Understand Urban Flow,” 2016.
- [18] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evskoff, and M. C. González, “The path most traveled: Travel demand estimation using big data resources,” *Transport Research Part C*, 2015.
- [19] C. M. Schneider, V. Beik, T. Couroneé, Z. Smoreda, and M. C. González, “Unraveling daily human motifs,” *J. R. Soc. Interface*, no. 10, May 2013.‘.
- [20] (03.06.2019) Teralytics. , s Interneta, <https://thenewstack.io/teralytics-takes-big-data-approach-human-movement/>
- [21] I. Stupar, P. Martinjak, V. Turk, and R. Filjar, “Socio-Economic Origin-Destination Matrix Derivation Through Contextualization of Material World,” *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2018.

Bibliografija

- [22] M. Filić, R. Filjar, and K. Vidović, “Graphical Representation of Origin-Destination Matrix in R Statistical Environment,” *36. skup o prometnim sustavima s međunarodnim sudjelovanjem AUTOMATIZACIJA U PROMETU*, November 2016.
- [23] Data description for urbancps. , s Interneta, <https://www-users.cs.umn.edu/~tianhe/BIGDATA/>
- [24] H. Zhao and Y. Zhao, “Traffic analysis zones - how do we move forward?” *AASHTO CTPP and TRB Joint Conference, Applying Census Data for Transportation*, 2017.
- [25] Z. Zheng, Z. Haung, F. Zhang, and P. Wang, “Understand coupling dynamic of public transport network,” *EPJ Data Science*, 2018.
- [26] E. Mellegard, “Obtaining origin-destination- matrices from cellular network data,” Master’s thesis, Chalmers University of Technology, Department of Mathematical Sciences, 2011.
- [27] A. Travassoli, A. Alsger, M. Hickman, and M. Meshbah, “How close the models are to the reality? Comparison of Transit Origin-Destination Estimates with Automatic Fare Collection Data,” *Australian Transport Research Forum*, 2016.
- [28] S. Bera and K. V. K. Rao, “Estimation of origin-destination matrix from traffic counts: the state of the art,” *European Transport Trasporti Europei n. 49*, 2011.
- [29] T. Djukić, “Reliability assessment of dynamic OD estimation methods based on structural similarity index,” *92nd meeting of the Transportation Research Board*, 2013.
- [30] T. van Vuren, “256 shades of gray - comparing OD matrices using image quality assessment techniques,” *Scottish Transport Applications Research (STAR)*, 2015.

Pojmovnik

ANPR Automatic Number Plate Recognition. 3

CDR Call Data Records ili Charging Data Records. 5–9, 12, 14, 19, 22

D4D Data for Development. 14

FCD Floating Car Data. 5

FPD Floating Phone Data (Floating Cellular Data). 5

GIS Geographic Information System. 9

GNSS Global Navigation Satellite System. 5, 7

GPS Global Positioning System. 8, 13

GSM Global System for Mobile (Communications). 5, 6

HW Home-Work. 7, 19

MSSIM Mean Structural Similarity index. 19, 20

ODM Origin-destination Martix. 1

OSM OpenStreetMap. 9, 11

POIs Points of Interest. 8, 9

POM Polazišno-Odredišna Matrica. 4, 6, 22

RSI Road Side Interview. 3

SSIM Structural Similarity index. 20

TAZ Traffic Analysis Zone. ix, 10, 11

WH Work-Home. 7, 19

Sažetak

Ovo je tekst u kojem se opiše sažetak vašega rada. Tekst treba imati duh rekapitulacije što je prikazano u radu, nakon čega slijedi 3-5 ključnih riječi (zamijenite dolje postavljene općenite predloške riječi nekim suvislim vlastitim ključnim riječima).

Ključne riječi — Polazišno-odredišna matrica, parametri kvalitete, usporedba

Abstract

This is a text where a brief summary of your work is outlined. The text should have a sense of recap of what was presented in the thesis, followed by 3-5 keywords (replace the general keyword templates below with some meaningful keywords of your own) .

Keywords —Origin-Destination Matrix, quality parameters, keyword 3

Dodatak A

Naslov priloga

A.1 Naslov sekcije

A.2 Naslov sekcije