

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

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

Rijeka, travanj 2019.

Vjera Turk
0069064924

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Diplomski rad

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

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

Rijeka, travanj 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, travanj 2019.

Ime Prezime

Zahvala

Zahvaljujem obitelji na podršci tijekom pisanja ovoga rada i korisnim raspravama i savjetima. Zahvaljujem xxxxx na podršku tijekom studiranja.

Sadržaj

Popis slika	ix
Popis tablica	x
1 Uvod	1
2 Izvorišno-odredišna matrica	2
2.1 Tradicionalni pristupi generiranju POM-a	2
2.1.1 Ankete	3
2.1.2 Prebrojavanje vozila	3
2.1.3 Modeliranje toka	3
2.1.4 Problematika i ograničenja tradicionalnih načina	3
2.2 POM iz zapisa o aktivnostima u javnoj pokretnoj mreži	4
2.2.1 Razlike u pristupima	4
2.2.2 Točnost položaja	5
2.2.3 Geometrija prostorne podjele	5
2.2.4 Dobre prakse u generiranju POM iz CDR	5
2.3 Drugi primjeri automatskog prikupljanja	6
2.3.1 Združena očitavanja prijamnika za satelitsku navigaciju	6
2.3.2 Javni prijevoz i <i>pametne kartice</i> (<i>Smart Card</i> sustavi)	6

3	Kontekstualizirane matrice	7
3.1	Kontekst iz samog izvora podataka o kretanju	7
3.2	Kontekst iz vanjskih izvora	7
3.2.1	<i>Points of Interes</i>	8
3.2.2	Open Street Map	8
3.3	Sredstvo (način) kretanja	8
4	Postojeće metrike za validaciju POM-e	9
4.1	Točnost POM-e	9
4.2	Metrike	9
4.3	Strukturalna sličnost	10
4.3.1	MSSI	10
5	Odnosni parametri kvalitete	12
5.1	Zajednički, objektivni kriteriji usporedbe	12
5.2	Komparacijski indikatori	12
5.2.1	Vremenski okvir	12
5.2.2	Razlučivost (Rezolucija)	13
5.2.3	Širina toka	13
5.2.4	Geometrija prostorne podjele	13
5.2.5	Definicija putovanja	13
5.2.6	13
5.2.7	Gustoća informacija - kontekst	13
5.3	Međuovisnost parametara	13
	Bibliografija	14
	Pojmovnik	15

Sadržaj

Sažetak	16
A Naslov priloga	17
A.1 Naslov sekcije	17
A.2 Naslov sekcije	17

Popis slika

Popis tablica

Poglavlje 1

Uvod

Izvorišno-Određišna matrica *eng. Origin-destination Matrix (ODM)* alat je koji omogućuje 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, 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.

Pojavljuje se potreba za objektivnom procjenom kvalitete ODM u odnosu na referentnu (kontrolnu). U ovom radu definirani su odnosni parametri kvalitete ODM te je razvijena metodologija usporedbe dviju ODM-a dobivenih različitim postupcima procjene i s podacima iz različitih izvora. Usporedba je obavljena korištenjem numeričkog i grafičkog oblika ODM.

Metodologija je izvedena u programskom okruženju za statističko računarstvo R te je demonstrirana njena primjena na slučaju usporedbe dviju ODM-a. Dobiveni rezultati komentirani su sa stajališta apsolutne i relativne točnosti matrica.

Poglavlje 2

Izvorišno-odredišna matrica

Izvorišno-odredišne matrica sadrži broj putovanja između svakog para zona unutar nekog područja za određeni vremenski okvir (u svakom smjeru zasebno). Svaka ćelija u matrici predstavlja broj putovanja iz izvorišne u ciljnu zonu.

Matrica može biti prikazana ukupno ili po pojedinim modulima prometnog sustava (osobna vozila, vozila javnog prijevoza, teretna vozila, pješaka itd). Može biti generirana za sadašnje ili planirano prometno opterećenje.

Iako intuitivno možemo zaključiti da se radi o kvadratnoj matrici, ove matrice su ponekad zbog ograničenja izvora ili modela predviđanja "rijetko popunjene", odnosno sadrže veliki broj nul-ćelija, pa su u nekim slučajevima prazni redovi ili stupci izbačeni, a posljedica toga je memorijski (i računski) manje zahtjevnja, preglednija ne-kvadratna matrica. Ova problematika istaknuta je u radu iz 2010. koji analizira kvalitetu matrica dobivenih iz anketa [5].

2.1 Tradicionalni pristupi generiranju POM-a

Traditional methods include running surveys within cities and estimating the flows between locations of the city from the feedback of those surveys. Such methods consume longer periods of time and are inaccurate at times. They usually span smaller population sample sizes and thus are more prone to biases.

Poglavlje 2. Izvorišno-odredišna matrica

Recent research in the domain of ubiquitous computing provided alternative methodologies for estimating more accurate ODs from user generated datasets like cell phone data.

2.1.1 Ankete

cijena anketiranja (u jednom od radova 10 eura po ispitaniku?) Road Side Interview (RSI)

2.1.2 Prebrojavanje vozila

ručno, video

2.1.3 Modeliranje toka

Mathematical modeling of traffic requires a lot of data and other information about the road network and the travel demand. The accuracy of the modeled traffic situation depends on the quality of the available information, and how this data is combined and weighted from different sources. The travel demand is a key component and nearly every traffic model requires a tableau OD/trip matrix/table specifying the travel demand between different places in the network. Such a tableau is called an Origin–Destination matrix, or OD-matrix for short; synonymously used terms are trip table or (origin–destination) trip matrix.

it is typically assumed that the travel demand is equal in both direction for all OD-pairs, [1]

2.1.4 Problematika i ograničenja tradicionalnih načina

Zastarijevanje

2.2 POM iz zapisa o aktivnostima u javnoj pokretnoj mreži

Today, with the ubiquity and pervasiveness of technology, data generated from mobile phones enable data analysts to better understand the behavior of individuals across many dimensions including their mobility patterns.

Unlike traditional analyses, the nature of this data mining approach forces us to first provide rigid, formal definitions of exactly what we mean by the terms origin, destination and journey. Here, both origins and destinations are subcategories of the overarching concept of a ‘stop’. A stop is defined as a set of contiguous network events that occur at the same location, over a minimum period of time.

This notion is parameterized to ensure we have sufficient confidence that any stop we have detected is not a transient location, but actually a location that the individual has actually settled in.

An algorithm must consequently be developed to exhaustively mine each person’s event series for such stops. Once achieved, the algorithm must next detect pairs of contiguous stops which occur at different locations and hence reveal movement. This pair can then be designated as a journey - the initiating stop becoming the journey’s origin, and the concluding stop as the journey’s destination.

coordination with traditional techniques is key to providing optimal solution in future.[2]

Global System for Mobile (Communications) (GSM)

2.2.1 Razlike u pristupima

Tranzitna t-POM,

Definiranje putovanja Period - Departure/Arrival time ”Flow between O-D pair w that departed its origin during time interval k” [3]

Grad, Država 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.

2.2.2 Točnost položaja

točnost položaja - aproksimacija s BS ili signal strenght (RSSI) + neka ona drugo
baza LTO(?)

2.2.3 Geometrija prostorne podjele

Heksagoni

Voronoi teslacije

Jedinice samouprave

Pravokutna mreža

2.2.4 Dobre prakse u generiranju POM iz CDR

Skaliranje CDR POM (primjerak – > pouplacija) (linking to transport infrastruc-
ture?) k-anonymization

2.3 Drugi primjeri automatskog prikupljanja

2.3.1 Združena očitavanja prijamnika za satelitsku navigaciju

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

82% putovanja javnim prijevozom naprave korisnici javnog prijevoza sa pametnim karticama. [4]

Poglavlje 3

Kontekstualizirane matrice

3.1 Kontekst iz samog izvora podataka o kretanju

Kontekst izvučen iz CDR (long term modeli- CDR više mjeseci/tjedana) HBW,HBO,NHB (The path most Traveled, HBW,WBH,HBO,NBO (Best Practices), HWO (?), HWH (Estimating origin-Destination flows using opportunistically collected mobile phone location data from one milion users in Boston Metropolitan Area) (2012 Estimation of urban commuting patterns using cellphone network data) - "MODE" kao kontekst, car/public transport/flight Terralytics

3.2 Kontekst iz vanjskih izvora

Some studies combined human mobility with land use or POIs data to segment districts in urban areas according to their functions or use. The type of data used to capture human mobility behavior varies between individual GPS traces [10, 11], taxi pick up/drop off locations as in [7, 12] , Call Detail Records (CDRs) as in [2, 8], social media check ins as in [13–15], and bus smart card data as in [16].

3.2.1 *Points of Interes*

Points of Interest (POIs) Points of Interes (Urban Attractors ima 22 kategorije koristi bazu s POI)

3.2.2 Open Street Map

Model raspodjele toka (naš model)

Izvor infrastrukture

OSM kao izvor infrastrukture (jedan od radova koristi za broj traka na auto cesti jedan za provjeru naseljenosti područja/broj katova zgrada)

3.3 Sredstvo (način) kretanja

Poglavlje 4

Postojeće metrike za validaciju POM-e

4.1 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 .

4.2 Metrike

Za procjenu kvalitete matrica dobivenih isključivo anketranjem u radu [5] 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.

[3] navodi 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). [6]

U [7] 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 Home-Work (HW) i Work-Home (WH) matrice dobivene iz Call Data Records ili Charging Data Records (CDR) u usporedbi s referentnim matricama dobivenim anketiranjem.

Travassoli u svome radu [4] 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 [8] za procjenu sličnosti matrica dobivenih iz CDR sa tada aktualnim matricama dobivenim anketiranjem.

4.3 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.

4.3.1 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

Poglavlje 4. Postojeće metrike za validaciju POM-e

prometu ideja o korištenju MSSIM za mjerenje sličnosti matrica se prvi puta spominje i demonstrira na simuliranim matricama dobivenim iz referentne matrice dodavanjem šuma. [6]

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

[6]

poboljšani

[9][10]

Poglavlje 5

Odnosni parametri kvalitete

Parametar se definira tako i tako

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.

5.1 Zajednički, objektivni kriteriji usporedbe

Isti grad

Isto doba godine

Isto vremensko razdoblje samo iz duljeg vremena moguće HW modele

5.2 Komparacijski indikatori

5.2.1 Vremenski okvir

Departure/Arrival time

5.2.2 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). točnost položaja

5.2.3 Širina toka

Ukupan broj odlazaka/dolazaka po vremenskom okviru za cijelu matricu

5.2.4 Geometrija prostorne podjele

(ne)uniformna podjela

5.2.5 Definicija putovanja

5.2.6

Infrastruktura

Sredstvo kretanja

5.2.7 Gustoća informacija - kontekst

5.3 Međuovisnost parametara

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

Bibliografija

- [1] A. Peterson, “The origin-destination matrix estimation problem- analysis and computations,” Ph.D. dissertation, Linköping Studies in Science and Technology, 2007.
- [2] J. Goulding, *Best Practices and Methodology for OD Matrix Creation from CDR data*, N/LAB, University of Nottingham, 2016.
- [3] 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.
- [4] A. Travassoli, A. Alsger, M. Hickman, and M. Meshbah, “How close the models are to the reality? comparison of transit oorigin-destination estimates with automatic fare collection data,” *Australian Transport Research Forum*, 2016.
- [5] 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.
- [6] T. Djukić, “Reliability assessment of dynamic od estimation methods based on structural similarity index,” *92nd meeting of the Transportation Reasearch Board*, 2013.
- [7] V. Frías-Martínez, E. Frías-Martínez, and C. S. Ruiz, “Estimation of urban commuting patterns using cellphone network data,” 2012.
- [8] E. Graells-Garrido and D. Saez-Trumper, “A day of your days: Estimating individual daily journeys using mobile data to understand urban flow,” 2016.
- [9] 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.
- [10] T. van Vuren, “256 shades of gray -comparing od matrices using image quality assessment tehniques,” *Scottish Transport Applications Research (STAR)*, 2015.

Pojmovnik

CDR Call Data Records ili Charging Data Records. 10

GSM Global System for Mobile (Communications). 4

HW Home-Work. 10

MSSIM Mean Structural Similarity index. 10, 11

ODM Origin-destination Martix. 1

POIs Points of Interest. 8

RSI Road Side Interview. 3

SSIM Structural Similarity index. 11

WH Work-Home. 10

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