LBS2022

17th Conference on Location Based Services

Proceedings of the

17th International Conference on Location-Based Services

Jukka M. Krisp, Liqiu Meng, Holger Kumke and Haosheng Huang (Editors) -

Selected accepted full papers are published in the special issue "Krisp, J. M., Meng L., Kumke H., Huang H. - Advances in Location Based Services http://dx.doi.org/10.1080/17489725.2023.2265867"

- Sara Irina Fabrikant (2023) Neuroadaptive LBS: towards human-, context-, and task-adaptive mobile geographic information displays to support spatial learning for pedestrian navigation, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2258100

- Xu Feng, Khuong an Nguyen & Zhiyuan Luo (2023) WiFi round-trip time (RTT) fingerprinting: an analysis of the properties and the performance in non-line-of-sight environments, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2239748

- Kenzo Milleville, Samuel Van Ackere, Jana Verdoodt, Steven Verstockt, Philippe De Maeyer & Nico Van de Weghe (2023) Exploring the potential of social media to study environmental topics and natural disasters, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2238663

- Nina Wiedemann, Henry Martin, Esra Suel, Ye Hong & Yanan Xin (2023) Influence of tracking duration on the privacy of individual mobility graphs, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2239190

- Johanna Vogt, Mario Ilic & Klaus Bogenberger (2023) A mobile mapping solution for VRU Infrastructure monitoring via low-cost LiDAR-sensors, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2238660

- Rui Li (2023) Augmented reality landmarks on windshield and their effects on the acquisition of spatial knowledge in autonomous vehicles, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2238661

- Irma Kveladze, Marina Georgati, Carsten Kessler & Henning Sten Hansen (2023) Analytics of historical human migration patterns: use cases of Amsterdam and Copenhagen, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2238658

Editors

Prof. Dr. Jukka M. Krisp, Professor of Applied Geoinformatics, University of Augsburg, Institute of Geography, Alter Postweg 118, 86159 Augsburg, Germany

Prof. Dr. Liqiu Meng, Professor of Cartography and Visual Analytics, Technical University of Munich, Arcisstrasse 21, 80333 Munich ,Germany

Dr. Holger Kumke, Researcher at Cartography and Visual Analytics, Technical University of Munich, Arcisstrasse 21, 80333 Munich ,Germany

Prof. Dr. Haosheng Huang, Professor of GIScience and Cartography, Ghent University, Department of Geography, Krijgslaan 281, S8, 9000 Gent, Belgium

This document contains the proceedings of the 17th International Conference on Location Based Services (LBS 2022), held on September 12-14, 2022 in Munich/Augsburg, Germany. All papers and abstracts have been accepted after a double-blind reviewing process.

The conference was organized by the ICA Commission on Location-Based Services, as a tandem conference including the teams of Applied Geoinformatics, (Prof. Krisp) University of Augsburg and the Cartography and Visual Analytics (Prof. Meng), Technical University of Munich, with the endorsement of the following organizations: International Cartographic Association (ICA) and Deutsche Gesellschaft für Kartographie (DGfK).

The editors would like to thank all the scientific committee members of LBS 2022 for reviewing the original submissions as well as Ms Verena Klasen with her help for compiling the proceedings.

Content

Session 1: privacy, social media & app usage

Session 2: wayfinding, simulation & human movement

Session 3: positioning and data acquisition

Session 4: wayfinding & navigation (indoor/outdoor) - work in progress

Session 5: privacy issues, ethics & uncertainty in data - work in progress

Session 6: visualization techniques for LBS and context - work in progress

Session 7: spatio-temporal data acquisition, processing & analysis I - work in progress

Session 8: spatio-temporal data acquisition, processing & analysis II - work in progress

Towards a GeoSocial Landmark Identification Model

MORITZ MÜHLMEIER, EVA NUHN & SABINE TIMPF

Geoinformatics Group, University of Augsburg, 86159 Augsburg, Germany

Tel: +49 821 598 2281 • E-Mail: eva.nuhn@geo.uni-augsburg.de

Keywords: Landmark Identification, Social Dimension, GeoSocial Model

Summary: *Research in human wayfinding shows, that integrating landmarks in route descriptions increases the success rate of navigational tasks for pedestrians. The salience of such landmarks is commonly measured using so called landmark dimensions. However, data collection for their attributes is difficult and time-consuming. A new promising data source emerged with the rise of geolocated social media content. We present a model to identify landmarks based on a social dimension using this content. We calculate a GeoSocial Score of objects in Augsburg using measures harvested from geosocial data and compare the outcomes with results of a survey. We conclude that geosocial data represent a reliable source of information to identify landmarks for pedestrians.*

Introduction

Landmarks are important elements for the communication of route descriptions, the orientation in, and navigation through space (Lynch, 1960; Allen, 2000; Michon and Denis, 2001). Most authors dealing with landmark identification build on the definitions of Sorrows & Hirtle (1999) and Raubal & Winter (2002) for landmark dimensions. These are the visual, the semantic, and the structural dimension. Other approaches focus on perceptual, cognitive, and contextual dimensions to model landmark salience (Caduff & Timpf, 2008). However, all of these approaches have in common that they need a wealth of different data sources to collect the information for all the attributes of these dimensions (Nuhn & Timpf, 2017). Due to the lack of data density, landmarks have hardly been picked up in actual, running navigation systems for pedestrians (Richter, 2017). The only service so far offering landmark-based verbal instructions is Whereis (Duckham et al., 2010). The underlying approach uses categories, which requires only data of an object's type and geographic location to determine an object's suitability as a landmark. However, this approach is based on the exploitation of points of interest (POIs) while landmarks are not limited to POIs (Richter & Winter, 2014).

These drawbacks can be overcome by using the social dimension, which describes "the way an object is practiced and recognised by a person or a group of people" (Quesnot & Roche, 2014 (p.1)). Since we are living in a 'geo-data-rich society' (Boulos, 2005), geospatial information to feed the social dimension is more accessible than ever. Volunteered Geographic Information that is produced by a large number of private citizens (Goodchild, 2007) includes data collected in social web platforms, such as Google place types (Google Place Types, 2022) and Foursquare (Foursquare, 2022). Quesnot & Roche (2014) argue that geosocial data represent a reliable source of information to precisely measure landmark semantic salience in an urban area. Their approach is based on Social Location Sharing, which consists of a check-in, which claims "I am/was at that place". Quesnot & Roche (2014) do not include Google Places API because it does not provide the information about check-ins. However, the Google place type, which describes an objects function, is a valuable source to identify social landmarks. We base our calculations on the user-generated Google place database, which is regularly updated by internet users (Quesnot & Roche, 2014). Furthermore, we consider Foursquare data, since they are a valuable source for the extraction of attributes regarding social prominence. We develop a GeoSocial model considering a social dimension and argue, that the model fed with data from Google place types and Foursquare is capable to identify landmarks, which would also be selected by humans. We apply the model in a pedestrian navigation scenario, where landmarks should be identified to be included in route directions. Finally, we evaluate the model by comparing the outcomes to the results of a survey.

GeoSocial Model – Basics

Our model considers the attributes place type, uniqueness, social prominence, and social activity to quantify the social dimension. We assign salience values to each place, based on these attributes and calculate a GeoSocial Score (GSS) to quantify the social salience of an object at a decision point (DP).

Place Type

We extract place types from Google Places database within a 50 meter radius around a DP. We assign all these places a salience value, based on the object's category of place type. Rousell and Zipf (2017) derive OSM place types, which are stored in each features attribute table, and reclassify them into broader, more general place type categories. They assign a weight value to each category, based on previous work of Duckham et al. (2010). We adapt their categorisation and transfer it to Google place types (Tab. 1). We introduce the weight values as tweight in our model.

Tab. 1: Place type category weight system.

Category	**tweight**	**Google Place Types**
shopping	0.8	clothing store, drugstore, jewelry store
grocery	0.8	supermarket, grocery store
gastronomy	0.7	café, bar, restaurant, bakery
health	0.5	doctor, dentist, pharmacy
office	0.5	insurance agency, lawyer, government office
service	0.5	hair care, travel agency, bank
transportation	1.0	rail station, transit station
religion	1.0	church, place of worship
leisure	1.0	park, plaza, sport facilities
tourist attraction	1.0	fountain, monument, theater

Uniqueness

Uniqueness investigates places and objects, where their associated function stands out in contrast to nearby objects (Quesnot and Roche, 2014). Following Rousell and Zipf (2017) and Quesnot and Roche (2014), we calculate the uniqueness score of an object as the ratio between the amount of places with the same place type (LM_{type}) and the total amount of objects in a 50 m radius at a single DP (LM_{total}). The result of this is subtracted from 1. In order to utilise the uniqueness metric in the GeoSocial Score as a weight multiplier, we apply a normalisation function to re-scale all values into a new range of 0.7 - 1.0. We choose the lower bound of uniqueness higher than 0 because of the multiplication of factors. We set the lower bound to 0.7 to avoid an overly low GSS which would result in case we would, e.g., use a lower bound of 0.001. Highly unique candidates get a unique value close to 1, while less unique places are close to a value of 0.7 (Eq. 1).

$$unique = 1 - (LM_{type}/LM_{total}) \rightarrow norm_{0.7\text{-}1.0}. \qquad (1)$$

Social Prominence

Bernardini and Peeples (2015) describe prominence as the 'Viewership' of elements in the landscape, in other words, the total number of viewers. In case an environmental feature has a great viewership, it is referenced a lot of times and thus perceived as salient. Each registered Google place can be reviewed by rating the place and leaving a written review. We apply a normalisation method per set of objects at a DP. The most reviewed place at a DP gets a weight value of 1.0, while the lowest rated place gets a weight value of 0.5 (Eq.

2). This means that the normalisation range for social prominence is with 0.5 higher than the range for uniqueness, since we consider social prominence as more important in this work.

$$\text{prominence} = LM_{review} \rightarrow norm_{0.5\text{-}1.0}. \qquad (2)$$

Social Activity

The two indicators derived from Foursquare are the amount of likes and the number of uploaded photos of a place. These indicators reflect the social activity of objects (Eq. 3). At each DP the range of objects' social activity score is re-scaled into a range of 1.0 for the lowest values and 1.3 for the highest values. This is due to the fact, that some places do not have any Foursquare data, namely no measurable social activity. This way, they remain unchanged when multiplied by a 1.0 weight but are normalised to the same range size (0.3) as uniqueness.

$$\text{activity} = LM_{likes} + LM_{photos} \rightarrow norm_{1.0\text{-}1.3}. \qquad (3)$$

GSS

We multiply the weights for the place type, the uniqueness, as well as the social prominence and activity to obtain the GSS (Eq. 4) for each object at a DP.

$$GSS = \text{tweight} * \text{uniq} * \text{prominence} * \text{activity} \qquad (4)$$

Eq. 5 shows the calculation steps for a 'Starbucks' café. We apply a tweight = 0.7 for the place type category gastronomy (Tab. 1). There are 5 other objects at the DP, making the café the least unique place type (uniqueness = 0.7). The 'Starbucks' café has 510 reviews (prominence = 0.61) and 23 likes and fotos (activity = 1.018). The resulting GSS is 0.30.

$$GSS_{Starbucks} = 0.7 * 0.7 * 0.61 * 1.02 = 0.30 \qquad (5)$$

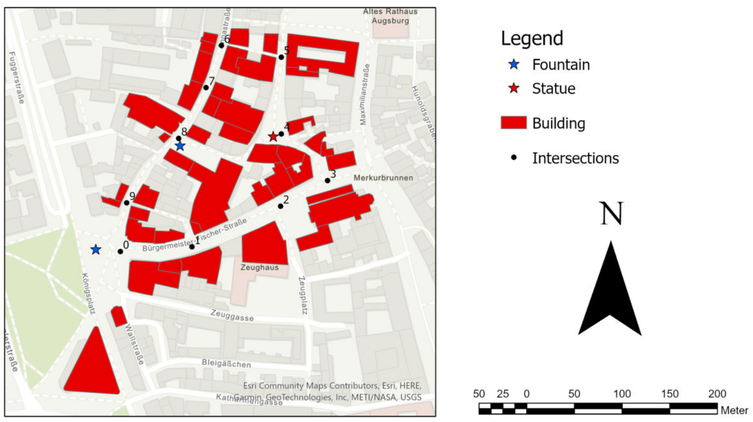


Fig. 1: Investigation area.

GeoSocial Model – Application

In this chapter we demonstrate the GeoSocial Model. We use a part of the inner-city of Augsburg as investigation area (Fig. 1). We harvested the data for this study between September and October 2021 from Google Places and Foursquare.

Place Type and Uniqueness

We identify 116 places in the investigation area. The place type and the resulting place type uniqueness are the first parameters in the GSS. With a total amount of 473 tags, each object has 4 tags on average. We classify the place types into the place type categories (Tab. 1). Fig. 2 shows the overall distribution of place type categories. The high number of shopping (47) and gastronomy (25) places is typical for pedestrian downtown areas.

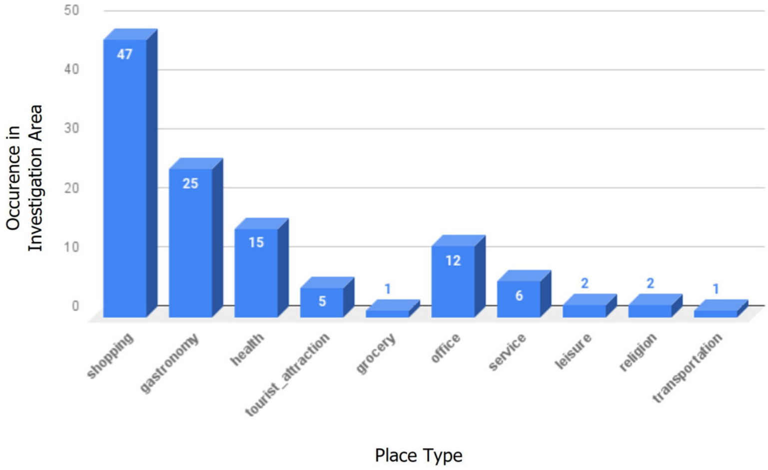


Fig. 2: Place type category in the investigation area.

Social Prominence

In order to eliminate noise that derives from Google Places, all places in the investigation area with less than 10 reviews are not used for further analysis. In total, 16904 reviews were extracted. Tab. 2 shows the most reviewed places on Google by place type category, highlighting the best place per place type. The 'McDonalds' is the most prominent place in the investigation area followed by the 'Starbucks' and 'Dunkin' Donuts'. All three of them are listed in the leading quick service restaurant companies of Germany by 2019, with the same hierarchical relationship (Statista, 2019). This may indicate that the social prominence of gastronomy places is linked to the quantitative popularity in Germany.

Tab. 2: Social prominence - most reviewed places by type.

Name	**Category**	**Reviews**
McDonald's Restaurant	Gastronomy	2376
Thalia	Shopping	1577
Travel agency	Service	523
REWE	Grocery	479
St. Anne's Church	Religion	295
Weberhaus	Tourist Attraction	204
Königsplatz Parc	Leisure	187
OZA	Health	142
Stadtwerke Customer Centre	Office	108
Moritzplatz	Transportation	22

The three most reviewed shopping places are the 'Thalia' bookstore (1577 reviews), 'SCHMID' (1128) clothing store, and the 'o2 Shop' (898). All of these mentioned places are chains that are present in several cities. This may also indicate that the quantitative popularity in Germany is linked to the local social prominence in the investigation area. More regional and local points-of-interest like St. Anne's Church (295), Weberhaus (204), or Moritzplatz (22) tend to have lower social prominence values.

Social Activity

Out of 116 places in the investigation area, 31 places show Social Activity derived from Foursquare database. In other words: only 26.72% of Google Places have associated Foursquare activity data. All places with type tourist_attraction in the investigation area show social activity. These places have on average 4.8 uploaded photos. Most photos are uploaded for the Fuggerdenkmal (Fig. 5). The other place types show similar patterns, with gastronomy and shopping covering 67.7% of all social activities.

GSS

Fig. 3 shows a scatter plot of the GSS for all 116 places. It is visible that most shopping places show low scores, with a few exceptions. These exceptions are prominent and draw great social significance. Gastronomy places scatter the most with diverse GSSs across the whole scale. Health, office, and service places reveal to be not socially active. The remaining place types tend to achieve higher scores of the GSS (tourist_attraction, grocery, leisure, and religion).

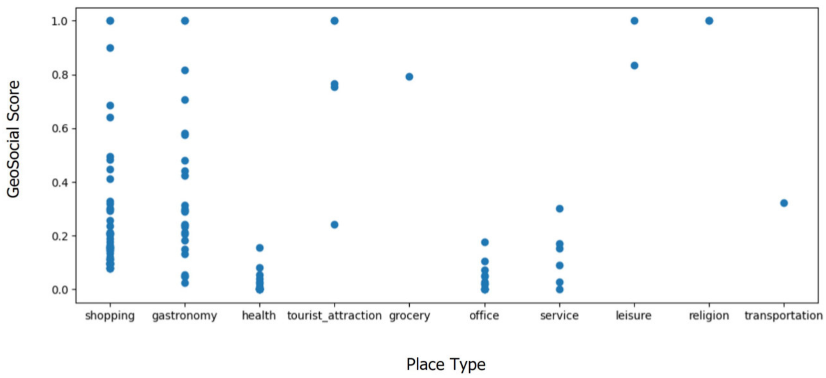


Fig. 3: GSS by place type.

GeoSocial Model – Evaluation

We compare the results of the GeoSocial model to the results of a survey. 51 participants selected at DPs in the inner-city (Fig. 1) objects useable as a landmark (Nuhn, 2020). We assign the number how often it was selected as a landmark to each object in the investigation area and compare this metric to the normalised GSS (highest GeoSocial Score at DP is 100%, the lowest is 0%). Then, we calculate a Pearson Correlation Coefficient. The coefficient indicates a moderate correlation (0.613) (Asuero et al., 2006). This suggests, that the GSS outputs similar objects as landmarks as the survey participants choose, but some adaptations might be needed in future work (see Conclusion and Outlook). Tab. 3 – Tab. 5 show the results of the comparison for 3 selected DPs (Fig. 4 – Fig. 6).

- DP 3: The model and the survey participants identify both the cultural building 'Weberhaus' as most prominent landmark (Tab. 3, Fig. 4). There is one building, which hosts multiple places belonging to shopping, gastronomy, and transportation (Fig. 4, Moritzplatz). We select the place with the highest score as a representative for that building polygon since most people associate one specific function with a building and often may not recognise multiple functions of a building.
- DP 4: The tourist attraction Fuggerdenkmal achieved the highest GSS and is selected the most from the survey participants (Tab. 4, Fig. 5). One building has no Google place type

although it hosts a museum. We extracted place types within a 50 meter radius around a DP. However, the Google place of the museum is not located within the radius. This is because the place types are located at the centroid of the polygons. Thus, in this case, the GSS cannot be calculated.

- DP 8: St. Anne's Church reaches the highest GSSs (Tab. 5, Fig. 6). It is not accessible from the DP and, additionally, is located behind a wall (Fig. 6). We assume that reviews, likes, and photos have been taken from people who entered the church from a DP on the other site of the church. The survey participants did not select St. Anne's Church but "Dr Scherer" as the most outstanding landmark. We believe that the participants did not select this landmark because of its health category but because of its function as a bank (Kreissparkasse) which is recognisable by an explicit mark. One building is missing in both, Google Places and Foursquare, since it was neither reviewed nor liked. Thus, the GSS cannot be calculated.

Based on these findings we can confirm that the model fed with data from Google place types and Foursquare is capable to identify landmarks, which would also be selected by humans. However, there might be adjustments necessary to improve the Geoscoial Score, which are discussed in the next Section.

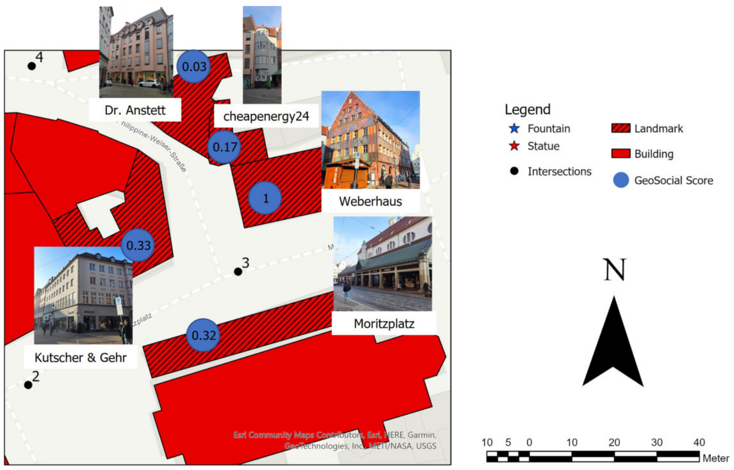


Fig. 4: DP 3.

Tab. 3: GSS DP 3.

Name	Category	Prominence	Activity	GSS	Landmark Selection
Weberhaus	Tourist Attraction	204	9	1	42
Kutscher + Gehr	Shopping	129	0	0.33	0
Moritzplatz	Transportation	22	0	0.32	5
cheapenergy24	Service	132	0	0.17	3
Dr. Anstett	Health	38	0	0.03	1

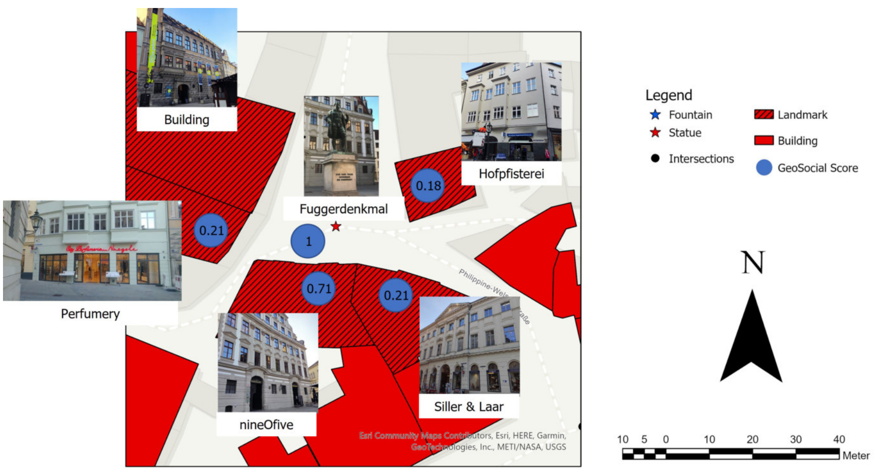


Fig. 5: DP 4.

Tab. 4: GSS DP 4.

Name	Category	Prominence	Activity	GSS	Landmark Selection
Fuggerdenkmal	Tourist Attraction	143	14	1	39
nineOfive	Gastronomy	270	3	0.71	6
Siller&Laar	Shopping	70	3	0.21	0
Perfumery	Shopping	71	2	0.21	0
Hofpfisterei	Gastronomy	17	0	0.18	0
Building (Museum)	-	-	-	-	6

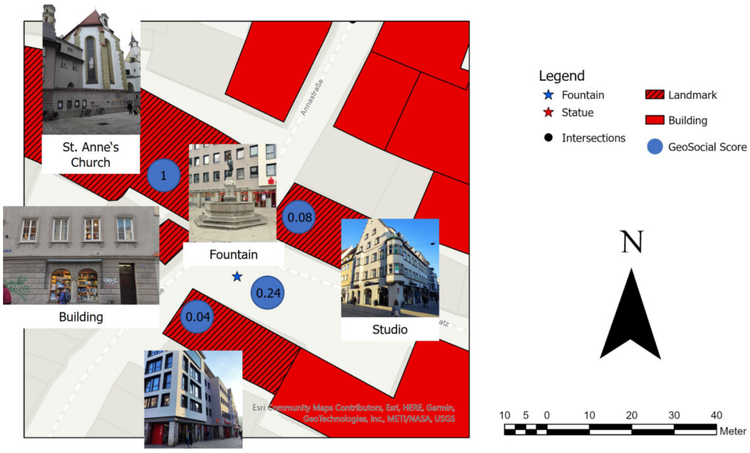


Fig. 6: DP 8.

Tab. 5: GSS DP 8.

Name	Category	Prominence	Activity	GSS	Landmark Selection
St. Anne's Church	Religion	295	23	1	9
Fountain	Tourist Attraction	14	1	0.24	15
Studio	Shopping	43	0	0.08	1
Dr. Scherer	Health	48	0	0.04	23
Building	-	-	-	-	3

Conclusion and Outlook

The Geosocial data, Google and Foursquare, represent a reliable source of information to identify landmarks for pedestrians. However, several problems still need to be addressed in future work. Our attributes are scaled only locally, that is there is always an object with the maximum GSS for each DP, independent of the absolute numbers of, e.g., views, likes, or photos. Thus, an alternative might be a global GSS for all the decision points. However, then, we need to find a solution for DPs where no landmark is identified. Additionally, recency of likes and reviews could be considered in future adaptations of the GSS. For example, a place with more current ratings could be more prominent than one with older ratings. Furthermore, we noticed that sometimes no reviews, likes, or photos are available for a specific object (compare DP 8). Moreover, sometimes not the landmark with the highest GSS, but another landmark seems more important to humans (compare DP 8). Additionally, landmarks with a high GSS might be located at a street intersection but not identified as most important for participants, since they are not accessible from the DP (compare DP 8). Furthermore, the location of the place type might not fall in a 50 meter radius around a DP (compare DP 4), although it could be an important landmark at the DP. The landmark dimensions can consider attributes such as accessibility in the structural dimension and the availability of explicit marks in the semantic dimension. Thus, the combination of our social dimension with the conventional landmark dimensions, seems promising for future work.

References

ALLEN, G. L., 2000: Principles and practices for communicating route knowledge, Applied Cognitive Psychology Psychology, 14, 333–359, https://doi.org/10.1002/1099-0720(200007/08)14:4<333::AID-ACP655>3.0.CO;2-C.

ASUERO, A. G., SAYAGO, A., & GONZALEZ, A. G. (2006). The correlation coefficient: An overview. Critical reviews in analytical chemistry, 36(1), 41-59, https://doi.org/10.1080/10408340500526766

BERNARDINI, W. & PEEPLES, M., 2015: Sight Communities: The Social Significance of Shared Visual Landmarks. In: American Antiquity 80, pp. 215–235. https://doi.org/10.7183/0002-7316.80.2.215.

BOULOS, M. N. K., 2005: Web GIS in practice III: creating a simple interactive map of England's Strategic Health Authorities using Google Maps API, Google Earth KML, and MSN Virtual Earth Map Control, International Journal of Health Geographics 4(22), 8, https://doi.org/10.1186/1476-072X-4-22.

CADUFF, D. & TIMPF, S., 2008: On the assessment of landmark salience for human navigation, Cognitive Processing 9(4), 249–267, https://doi.org/10.1007/s10339-007-0199-2

DUCKHAM, M., WINTER, S. & ROBINSON, M., 2010: Including landmarks in routing instructions, Journal of Location Based Services 4(1), 28–52, https://doi.org/10.1080/17489721003785602.

FOURSQUARE, 2022: Foursquare. https://foursquare.com/products/places/, Accesed: February, 2022.

GOODCHILD, M. F., 2007: Citizens as sensors: the world of volunteered geography, GeoJournal 69(4), 211–221, https://doi.org/10.1007/s10708-007-9111-y.

GOOGLE PLACE TYPES, 2022: https://developers.google.com/maps/documentation/places/web-service/supported_types, Accessed: Febraury, 2022.

LYNCH, K., 1960: The image of the city, MIT press, https://doi.org/10.1177/000271626133600164.

MICHON, P.-E. & DENIS, M., 2001: When and why are visual landmarks used in giving directions?, In: Spatial Information Theory. COSIT 2001, edited by Montello, D. R., vol. 2205 of LNCS, pp. 292–305, Springer, https://doi.org/10.1007/3-540-45424-1_20.

NUHN, E. & TIMPF, S., 2017: A multidimensional model for selecting personalised landmarks, Journal of Location Based Services 11(3-4), 153–180, https://doi.org/10.1080/17489725.2017.1401129.

NUHN, EVA, 2020: Modelling of personalised landmarks. PhD thesis, University of Augsburg, Faculty of Applied Computer Science.

QUESNOT, T & ROCHE, S., 2014: Measure of Landmark Semantic Salience through Geosocial Data Streams. In: ISPRS International Journal of Geographic Information, pp. 1–31. https://doi.org/10.3390/ijgi4010001.

RAUBAL, M. & WINTER, S., 2002: Enriching Wayfinding Instructions with Local Landmarks, In: Geographic Information Science. GIScience 2002, edited by Egenhofer, M. J. and Mark, D. M., vol. 2478 of LNCS, pp. 243–259, Springer, https://doi.org/10.1007/3-540-45799-2_17.

RICHTER, K.F. & WINTER, S, 2014: *Landmark: GIScience for Intelligent Services*; Springer: Cham, Switzerland.

RICHTER, K.-F., 2017: Identifying landmark candidates beyond toy examples – a critical discussion and some way forward, KI-Künstliche Intelligenz 31(2), 135–139, https://doi.org/10.1007/s13218-016-0477-1.

ROUSELL, A. & ZIPF, A., 2017: Towards a landmark-based pedestrian navigation service using osm data', ISPRS International Journal of Geo-Information 6(3), 22, https://doi.org/10.3390/ijgi6030064.

SORROWS, M. E. & HIRTLE, S. C., 1999: The Nature of Landmarks for Real and Electronic Spaces, In: Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science. COSIT 1999, edited by Freksa, C. and Mark, D. M., vol. 1661 of LNCS, pp. 37–50, Springer, https://doi.org/10.1007/3-540-48384-5_3.

STATISTA, 2019: Leading quick service restaurant companies. https://www.statista.com/statistics/734333/germany-quick-service-restaurant-companies-by-outlets/ Accessed: February 2022.

Please find this paper in the following DOI

Xu Feng, Khuong an Nguyen & Zhiyuan Luo (2023) WiFi round-trip time (RTT) fingerprinting: an analysis of the properties and the performance in non-line-of-sight environments, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2239748

Please find this paper in the following DOI:

Johanna Vogt, Mario Ilic & Klaus Bogenberger (2023) A mobile mapping solution for VRU Infrastructure monitoring via low-cost LiDAR-sensors, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2238660

Exploration of User Privacy in 802.11 Probe Requests with MAC Address Randomization Using Temporal Pattern Analysis

TOMAS BRAVENEC, JOAQUÍN TORRES-SOSPEDRA, MICHAEL GOULD, TOMAS FRYZA

Institute of New Imaging Technologies • Universitat Jaume I • Avenida Sos Baynat, s/n • 12071, Castellón de la Plana, Spain

Department of Radio Electronics • Brno University of Technology • Technicka 12 • 61600, Brno, Czech Republic

Centro Algoritmi (CALG) • University of Minho • Av. da Universidade • 4800-058, Guimarães, Portugal

Tel.: (+34) 964 72 92 35 • E-Mail: bravenec@uji.es

Keywords: MAC randomization, temporal analysis, privacy, tracking, probe requests

Summary: *Wireless networks have become an integral part of our daily lives and lately there is increased concern about privacy and protecting the identity of individual users. In this paper we address the evolution of privacy measures in Wi-Fi probe request frames. We focus on the lack of privacy measures before the implementation of MAC Address Randomization, and on the way anti-tracking measures evolved throughout the last decade. We do not try to reverse MAC address randomization to get the real address of the device, but instead analyse the possibility of further tracking/localization without needing the real MAC address of the specific users. To gain better analysis results, we introduce temporal pattern matching approach to identification of devices using randomized MAC addresses.*

Introduction

The technology that changed our lives the most in last decade is without question the introduction of the smartphone. Having internet connection on our person as we move through the world has had major impact on both our professional and personal lives. The practically constant connection to the internet, be it through cellular data or Wi-Fi, introduces a question of how much privacy, locational and otherwise, are people giving up. They are often giving up their privacy willingly for the use of services that make their daily lives easier; in other cases they do not know who or what may be identifying and/or tracking them.

Devices using Wi-Fi (also called WLAN) for connecting to the internet are extremely common, with most people having at least one around them at all times, for example mobile phone, smartwatch, laptop, and smart TVs. Since the majority of these devices are connected to the internet through wireless networks, the issue of privacy and device tracking on those networks should be something people are aware of. Our devices are communicating with the surrounding world using standardized protocols. For instance, a device in a IEEE 802 network is uniquely identified by the Media Access Control (MAC) address, which is used in all the messages involving the device. The device probe request is a type of wireless frame used to gather information about Wi-Fi access points in the proximity of a device. This is beneficial to the users as the device can identify and connect to a known access point without any user input, to switch to anohter AP with better coverage in a large public Wi-Fi network, as well as help with increasing accuracy of geolocation navigation by checking nearby Wi-Fi devices and comparing the signal strenght of detected access points with previously detected ones at the same location. These probe requests can be a major weak point of a Wi-Fi protocol, since they allow for non-cooperative user tracking if the device does not use enough privacy measures such as MAC address randomization.

Tracking using Wi-Fi protocols can vary as they can be used to determine the past whereabouts of users, current presence, or both. The past locations of devices can be determined if the devices are transmitting the preferred network list (list of the networks the device was connected to in the past) which can be matched to location using access point databases like WIGLE (2022). The current presence tracking can be done using a fingerprinting approach or in the case of devices without randomized MAC addresses, just by matching the globally unique MAC addresses of separate probe requests.

At the time of writing, most of the major operating systems have implemented some kind of privacy protection measures, that are helping to protect users from non-cooperative tracking. But the implementations and efficiency of those measures vary.

In this paper we explore the current state of privacy related measures in probe requests. We analyse how the situation has changed since the period before MAC randomization was first introduced, and we propose additional measures to further increase privacy. The main contribution of this paper is in the new angle of analysing the probe request datasets from the temporal point of view. We present a temporal pattern matching approach to identifiying devices with randomized MAC addresses through the pattern of their appearances in time.

Related Work

The tracking of mobile device users using passive sniffing of probe requests has been a focus of research for quite a while now. For example Musa & Eriksson (2012) used probe requests for urban mobility tracking. The privacy vulnerability of the probe request frames was already proven in several publications by Ningning et al. (2013) or by Cunche et al. (2014) prior to the implementation of MAC address randomization. After the introduction of MAC address randomization in iOS 8 in 2014 (Vasilevsky et al., 2019), researchers worked to determine the inner working of the randomization technique Apple used (Freudiger, 2015). In other research the authors focused on determining the real MAC addresses assigned by each manufacturer (Martin et al., 2016). Di Luzio et al. (2016) determined the origin of people at large events using probe requests collected at 2 political events before elections in Italy and results were matching the official voting reports. Matte et al. (2016) provided details on bypassing MAC randomization with the use of temporal analysis, by exploiting the device specific timings between subsequent probe requests or scan instances. Martin et al. (2017) created very deep study of MAC address randomization and explored all the times it fails. Gu et. al. (2020) proposed an encryption for 802.11ac devices.

Current Implementation of MAC Randomization

Although the IEEE Standards Association Standards Board specified in 2018 a standard amendment 802.11aq-2018 (IEEE, 2018) considering randomization of MAC addresses, there is still no standard for actual implementation of randomization. This means each and every manufacturer and software developer can decide how to implement it in their own manner.

Addresses can be assigned either by the manufacturer or locally by the device network controller. The way the address is assigned is differentiated by the 2nd least significant bit of the first byte of the MAC address B1 as shown in Figure 1. If the bit is set, the MAC address was assigned locally. The least significant bit of the first byte B0 describes if the MAC address is unicast or multicast. For the majority of devices, this bit is set to 0. The unicast/multicast bit is set to 0 for individual devices and only set to 1 for device groups. This makes the distinction between globally unique and randomized MAC address of individual devices very simple as the 2nd digit of randomized MAC address in hexadecimal format can only be 2 (0010), 6 (0110), A (1010) or E (1110).

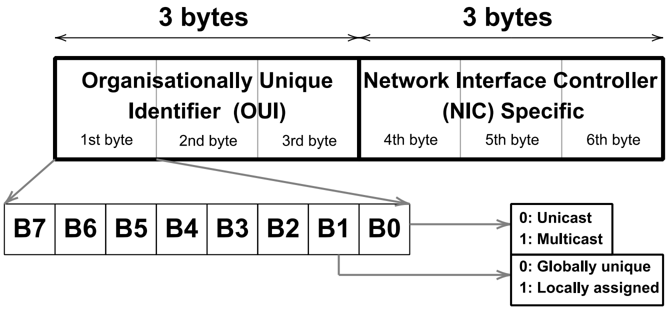


Fig. 1: Structure of MAC address with the functional bits

Dataset

We decided to collect a dataset to perform the analsysis. As we work with real data containing personal information, it stays to say that we only approached the analysis from the implementation of Wi-Fi protocol, specifically to explore potential privacy issues in current implementations. The data collection was limited to only passive capture and store of 802.11 management frames which we then anonymized before starting with analysis. The annonymization is done by hashing last 3 bytes of the MAC address, preffered network list, UUID-E and several other fields containing user specific information. From the anonymized data we do not observe personally identifiable information. Even though the collected data contains both real and randomized MAC addresses, it is not possible for us to match a MAC addresses to specific individuals because the analysis was done over annonymized version of the dataset, and we did not collect the data with presence in the office or the building.

To base our research work on up to date data, we collected probe requests at our small office for 6 days in December 2021. The office is in the corner of the 5th floor and during peak times is occupied by about 15 researchers.. In that time we collected 340,360 probe requests. The data was collected using an ESP32 micro controller with connected micro SD card for storage of the collected probe requests. The firmware created for the ESP32 micro controller to capture probe requests and save them in standardized way readable by the Wireshark and similar packet analysis tools is publicly available from a GitLab repository (Bravenec, 2021).

About 10% of captured probe requests contained WPS (Wi-Fi Protected Setup) sections, which provide additional information about the device, starting with device name, manufacturer, and model. Since many devices use the name of their owner, this may pose a privacy leak in devices transmitting this additional information which is unnecessary for correct functioning of probe request frames. Even sending a device manufacturer name can reveal the user identity if the device itself is less common than others (e.g. Susie is the only Motorola user here). The most important issues of this WPS section, though, is the presence of UUID-E (Universally Unique IDentifier-Enrollee) data which is unique for a device since it is acquired using the globally unique MAC address of the device and does not change. Devices transmitting probe requests containing UUID-E are then easily localized as their globally unique MAC address can be recovered using UUID-E reversal techniques - by looking up the globally unique MAC address from hash tables (Martin et al., 2016). Therefore, we have hashed that additional information provided in the WPS Section to ensure privacy of users.

Analysis

In the past, the tracking of mobile devices using only probe requests was not very difficult as there were several factors that made identification of a single device fairly straightforward. These include nonrandomized MAC addresses, consecutive Sequence Numbers, common time difference between 2 probe requests, or extended information in the Information Element like supported transfer rates and vendor information.

MAC addresses

Even though MAC addresses cannot be used effectively to locate most modern devices, they can still be used to identify a single device during a single scan. From the analysis of the data collected at our office, the devices do not randomize MAC addresses after every probe request. This makes identification of a single scan instance from one device very easy, since they keep the same address for the scanning sequence, or multiple sequences. A solution to increase privacy would be to randomize the MAC address for every probe request, or at least more often than the devices do at the moment.

Sequence Numbers

Sequence numbers in probe request packets allow for another opportunity to easily identify packets coming from a single device during one scan instance, without the need to check the MAC address. The reason for this is the incremental nature of sequence numbers in probe requests coming from a single device. Every time the device sends a packet, the sequence number increments by 1. The sequence number can increase by more than 1, which happens if a device sends another packet or frame between 2 subsequent probe requests.

Addressing this issue would be fairly straightforward by using random sequence number for each probe request. This, combined with randomization after every probe request, would make identification of packets coming from a single device a lot more challenging as new techniques for identification through probes would be required.

Fingerprinting with Information Elements

To identify devices we collect the device specific information fields available in probe requests, out of which we create a single unique identifier (Loh et al. 2008). The information element fields in probe requests can contain various additional data, starting with supported transfer speeds, information about the vendor of the wireless chip inside of the device, and including the connected peripherals and device name. As mentioned in the section describing the dataset, WPS information might also be present, which contains enough information to create a unique fingerprint of the device. The biggest issue there is the UUID-E field which is unique per device and makes MAC address randomization pointless in devices that transmit WPS data, since instead of MAC address the UUID-E can be used to identify a device. And that is without considering UUID-E reversal techniques which can be used to determine the globally unique MAC address of the device (Martin et al., 2016).

For fingerprint creation we use all of the fields that remain constant for one device between transmissions. Supported transmission speeds, vendor information, WPS field and others are used to create a hash using the SHA512 algorithm. This ensures we have an unique fingerprint for information element of all devices. All of the fields in our device fingerprint are presented in Table 1, with the frequency of occurrences in the data collected at our office. As can be seen, the supported data rates are presented in 100% of collected probe requests, with extended list of supported data rates being missing from just 0.05% of the probes. The HT Capabilities (802.11n specific information regarding supported frequency bandwidth etc.) were also present in a majority of probe requests, followed up by extended capabilities and at least 1 vendor specific field, though the most common number of vendor specific fields for one probe request in the data collected in our lab was 4, in about 30 % of all probe requests. Since devices from the same vendor will have the same vendor specific fields, having 4 vendor specific fields the same, increases the probability that the devices are the same.

Fingerprinting SSID lists

By using previously mentioned techniques, it is easy to differentiate all probe requests sent by a single device in a single scan instance. Knowing that all probe requests came from a single device then allowed us to list all the different SSIDs in those probe requests. By using sets with each SSID represented only once, we can use set similarity as in equation (1) to calculate a probability of two devices being in fact a single device by using the transmitted SSID list.

$$p = \frac{set(A)\ \textbf{and}\ set(B)}{set(A)\ \textbf{or}\ set(B)} \qquad (1)$$

There is also a possibility for the attacker to identify the users directly through the SSIDs from the preffered network list, as there is a possibility to match some of the networks directly to people (for example SSID of network at university in another country, while we know our collegue is the only one around that used to study there).

Information Element	Included in Probes	[%]
Supported Rates	340360	100.00
Extended Supported Rates	340198	99.95
HT Capabilities	312227	91.73
VHT Capabilities	20252	5.95
Extended Capabilities	232918	68.43
Vendor Specific Elements	194801	57.23
1 Vendor Specific Element	29681	8.72
2 Vendor Specific Element	47375	13.92
3 Vendor Specific Element	8661	2.54
4 Vendor Specific Element	104559	30.72
5 Vendor Specific Element	4525	1.33
WPS – UUID-E	35908	10.55
Total Collected Probe Requests	340360	

Table. 1: Probe request fields used to create device fingerprint and frequency of occurrence in data collected in our lab

Device identification

Combining the use of non-randomized MAC addresses, device fingerprint elements, use of transmitted SSIDs to differentiate devices, and UUID-E available in the probe requests with WPS field,we have enough information to identify a single Wi-Fi scan instance (Algorithm 1) as well as reappearance of a device. Even with MAC randomization, the information elements in the probe requests allow the adversary to identify devices.

After identifying the Wi-Fi scan instances, we start with device identification. Here we first check if the MAC addresses of 2 separate instances are the same. If they are we can consider the instances as the same device. If the MAC addresses are randomized or different from each other, we check for the presence of WPS fields, and subsequently check the UUIDE field and evaluate if the device is the same or not. In case the WPS field is not included and MAC addresses are not matching, we determine the similarity using the information elements section of probe requests and calculate a similarity score between the two preferred network lists. If the similarity is higher than a set threshold we can consider the scan instances to be from the same device. Since the preffered network list revealed through probe requests is in majority of the cases quite short and in many cases can be incomplete, the threshold was set to >0.5. Since with two transmitted SSIDs in each scan instance, one identical SSID will result in the similarity of 0.5. And since Wi-Fi networks have quite unique names, we take devices with at least two matching SSIDs as similar devices. The process of identifying a single device is shown in Algorithm 2.

Temporal Analysis

One of the more difficult parameters to mask for a single device sending multiple probe requests is the time difference between 2 probe requests. From our analysis of the Sapienza Probe Request dataset created by Barbera et. al (2013), slightly more than 98 % of subsequent probe requests sent by a single device are transmitted less than 65 milliseconds apart. These bursts of transmitted probe requests can be used for fingerprinting of the device. This is useful in conjunction with incrementing sequence numbers to distinguish two different devices and will be a potential threat to the users in the future, since the incrementing sequence number could reveal one device using multiple MAC addresses after every probe request.

We did not use the time difference between two probe requests, since devices do not change their MAC address during the scan instance. Instead we used different approach to time analysis. We used all of the similar device data that we got during device identification and analysed the recurrent appearances of each device and possible similarity to others. This way we discovered a pattern that allowed us to identify cases where single device looked like several devices. We did this by considering scan instance appearances of one device and clustering them together based on time. We then compared the number of clusters between devices. In case two devices had the same amount of appearance clusters, we checked the overlay between clusters. Subsequently we decided if the devices were in the end single device misidentified as many, or skip it it and move to the next device.

Algorithm 1 Scan Instance Identification

```
variables
    probe.mac, MAC address of the probe request
    probe.has_wps, Probe request with WPS field
    probe.uuide, UUID-E of the probe request
    probe.ie, Information Element of the probe request
    probe.sn, Sequence number of the probe request
end variables
if probe_1.mac = probe_2.mac then
    if probe_1.has_wps and probe_2.has_wps then
        if probe_1.uuide = probe_2.uuide then
            return True                         ▻ True - Same instance
        else
            return False                        ▻ False - Different instance
        end if
    end if
    if probe_1.ie = probe_2.ie then
        if probe_1.sn < probe_2.sn < probe_1.sn + 5 then
            return True                         ▻ True - Same instance
        else
            return False                        ▻ False - Different instance
        end if
    else
        return False                            ▻ False - Different instance
    end if
else
    return False                                ▻ False - Different instance
end if
```

Algorithm 2 Device Identification

```
variables
    instance.mac, MAC address of the instance
    instance.has_wps, Instance with WPS field
    instance.uuide, UUID-E of the instance
    instance.ie, Information Element of the instance
    instance.SSIDs, List of SSIDs from one instance
    threshold, Minimum similarity threshold
end variables
if instance_1.mac = instance_2.mac then
    return True                                 ▻ True - Same device
else if instance_1.has_wps and instance_2.has_wps then
    if instance_1.uuide = instance_2.uuide then
        return True                             ▻ True - Same device
    else
        return False                            ▻ False - Different device
    end if
else if instance_1.ie = instance_2.ie then
    p = (set(instance_1.SSIDs) and set(instance_2.SSIDs)) / (set(instance_1.SSIDs) or set(instance_2.SSIDs))
    if p > threshold then
        return True                             ▻ True - Same device
    else
        return False                            ▻ False - Different device
    end if
else
    return False                                ▻ False - Different device
end if
```

Results

From the 340,360 probe requests collected at our office, we identified in total 125,983 scan instances. After following Algorithm 2, we got 1023 devices as is represented in Figure 2a). As a single instance we count any single probe request or burst of probe requests according to Algorithm 1. These instances were then clustered based on their similarity following the Algorithm 2. This way we were able to match at least two instances to a single device. If the tested instance showed no similarity to others, that instance was discarded as a single instance device that we had no way to track or to locate.

For devices that do not randomize their MAC addresses, the tracking is very effective and we can easily see when the device was inside of our office or in its proximity. Reason being, the unique identifier is the MAC address, which never changed. Due to this we were able to identify a significant number of devices that could be easily tracked and analysed for presence patterns. Between those devices were also a few that never left the proximity of the probe request sniffer as well as some that showed up in monitored range only for a few minutes. As presented in Figure 2a), 212 devices did not use MAC randomization. As an example comparison of presence in time for 8 such devices is shown in Figure 3.a). As a single instance we count any single probe request or burst of probe requests according to Algorithm 1. These instances were then clustered based on their similarity following the Algorithm 2. This way we were able to match at least two instances to a single device. If the tested instance showed no similarity to others, that instance was discarded as a single instance device that we had no way to track or to locate.

For devices that do not randomize their MAC addresses, the tracking is very effective and we can easily see when the device was inside of our office or in its proximity. Reason being, the unique identifier is the MAC address, which never changed. Due to this we were able to identify a significant number of devices that could be easily tracked and analysed for presence patterns. Between those devices were also a few that never left the proximity of the probe request sniffer as well as some that showed up in monitored range only for a few minutes. As presented in Figure 2a), 212 devices did not use MAC randomization and the example of presence in time for 8 devices is shown in Figure 3.

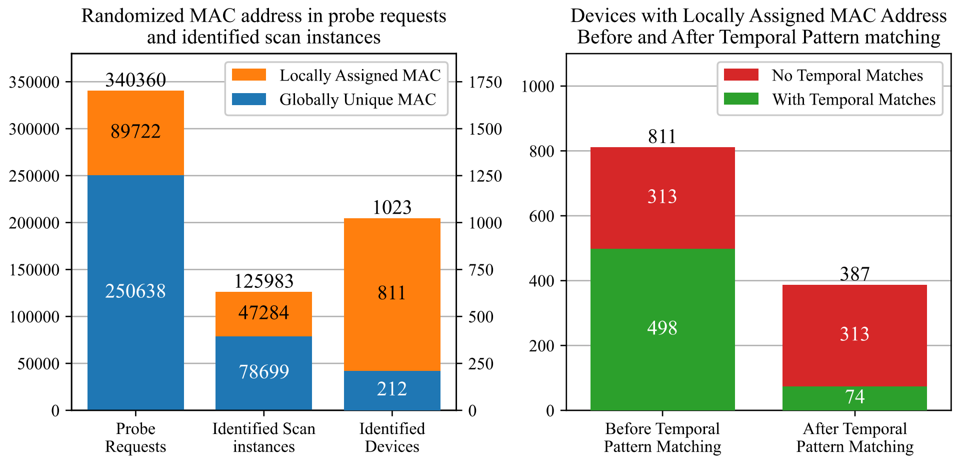


Fig. 2: Dataset information and device identification: a) Probe Requests, Identified Scan Instances and Identified Devices, b) Identified Devices before and after Temporal pattern Matching

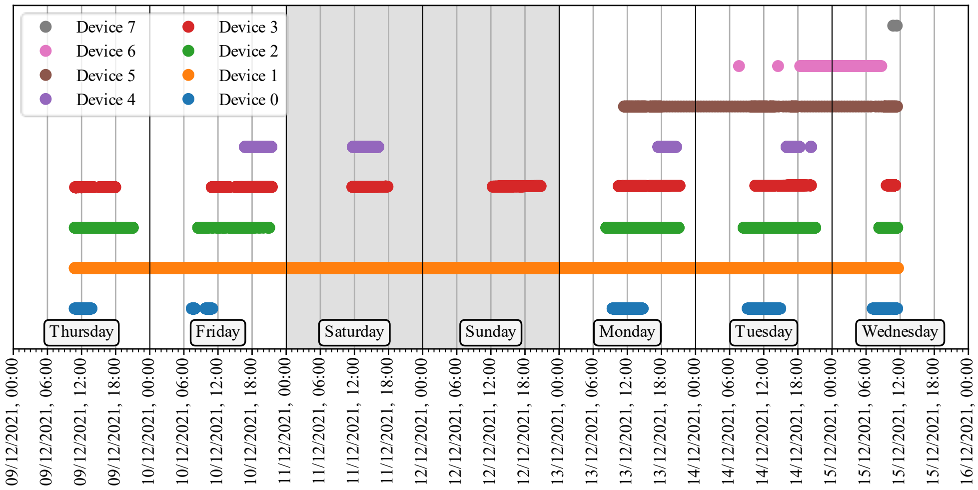


Fig. 3: Occurrence of several devices identified by the usage of globally unique MAC address

The identification of devices randomizing MAC addresses is more complicated, but despite MAC randomization making the process more difficult, we were able to identify many devices using the techniques mentioned before in the Analysis section. The results of our analysis can be seen on 8 devices using randomized MAC address in Figure 4. Even with the more complicated approach to identification, from the resulting data, the analysis of user presence is still possible.

The Algorithm 2 provides us with instances clustered as one device. The instance matching is not 100% accurate and in some cases, especially in those considering devices with randomized MAC addresses, can misidentify a single device as several devices. Using the Algorithm 2, we matched the 125,983 instances to 1023 devices, which can be seen in Figure 2a). After the instance matching, we used the temporal analysis we proposed on the identified devices. We managed to detect similarity in between 498 misidentified devices, and reduce this amount to only 74 devices using MAC randomization. 313 devices with locally assigned MAC addresses did not match temporal pattern of other devices. The amount of devices with locally assigned MAC address before and after temporal pattern matching can be seen in Figure 2b). The probe request transmission patterns were quite closely matching each other, as can be seen on Figure 5, which led us to identifying these appearances as a single device or single user carrying multiple devices.

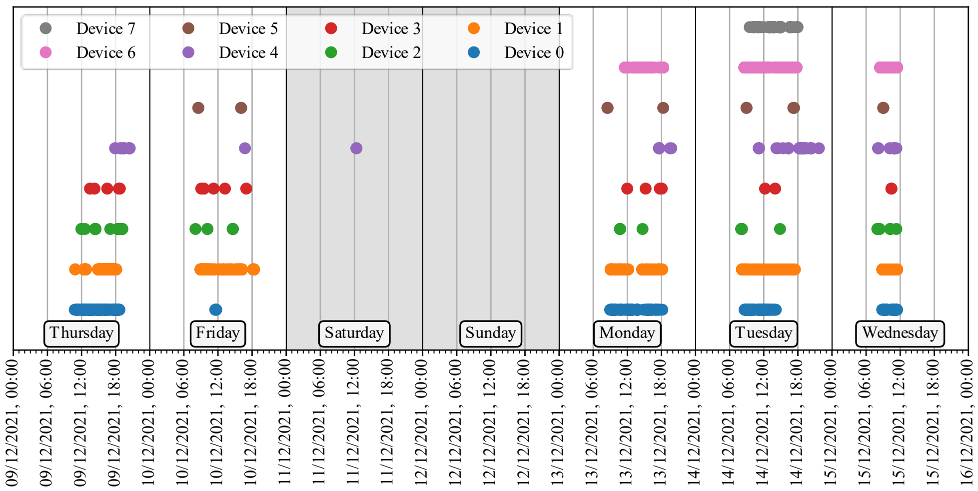


Fig. 4: Occurrence of several devices identified despite the use of MAC randomization

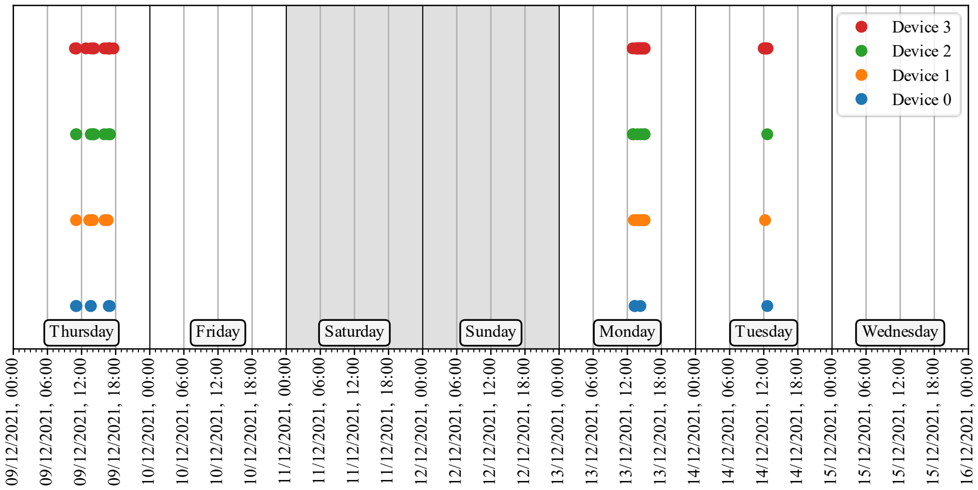


Fig. 5: Occurrence of single device misidentified as multiple devices, later identified as single device through the similarity in temporal patterns

Conclusions and Future Work

In this paper we explored the current state of privacy regarding probe requests in the 802.11 standard using our captured dataset of probe requests. Through deterministic methods we explored the possibilities to bypass and identify devices without the need for using the globally unique MAC address. From our results we managed to track many devices with and without locally assigned MAC addresses.

We also introduced an approach to use temporal pattern matching to identify device appearing as several devices due to the use of MAC address randomization. Using this trchnique we managed to reduce 498 identified devices to just 74. This makes the temporal pattern matching quite an effective technique for detecting devices despite using MAC address randomization.

For the future we plan to continue the exploration of privacy with probe requests and we plan to collect and publish a new probe request dataset. We also plan to release a small dataset without the use of anonymization techniques, from our controlled environment and with consent of everyone involved in the data capture.

Acknowledgement

The authors gratefully acknowledge funding from European Union’s Horizon 2020 Research and Innovation programme under the Marie Sklodowska Curie grant agreement No. 813278 (A-WEAR: A network for dynamic wearable applications with privacy constraints, http://www.a-wear.eu/) and No. 101023072 (ORIENTATE: Low-cost Reliable Indoor Positioning in Smart Factories, http://orientate.dsi.uminho.pt). This work does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

References

WIGLE, 2022. Wireless network mapping. https://www.wigle.net/

MUSA, A., & ERIKSSON, J., 2012. Tracking Unmodified Smartphones Using Wi-Fi Monitors. Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems (p. 14). Toronto, Ontario, Canada: Association for Computing Machinery.

NINGNING, C., WANG, X. O., WEI, C., MOHAPATRA, P., & SENEVIRATNE, A., 2013. Characterizing privacy leakage of public wifi networks for users on travel. 2013 Proceedings IEEE INFOCOM (p. 9). Turin, Italy: IEEE.

CUNCHE, M., KAAFAR, M.-A., & BORELI, R. (2014). Linking wireless devices using information contained in Wi-Fi probe requests. Pervasive and Mobile Computing, 56-69.

VASILEVSKI, I., BLAZHEVSKI, D., PACHOVSKI, V., & STOJMENOVSKA, I., 2019. Five Years Later: How Effective Is the MAC Randomization in Practice? The No-at-All Attack. ICT

Innovations 2019. Big Data Processing and Mining (pp. 52-64). Springer International Publishing

FREUDIGER, J., 2015,. How Talkative is Your Mobile Device? An Experimental Study of Wi-Fi Probe Requests. Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks (p. 6). New York, NY, USA: Association for Computing Machinery.

MARTIN, J., RYE, E., & BEVERLY, R., 2016. Decomposition of MAC Address Structure for Granular Device Inference. Proceedings of the 32nd Annual Conference on Computer Security Applications (pp. 78-88). Los Angeles, California, USA: Association for Computing Machinery.

DI LUZIO, A., MEI, A., & STEFA, J., 2016. Mind your probes: De-anonymization of large crowds through smartphone WiFi probe requests. IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications (p. 9). San Francisco, CA, USA: IEEE.

MARTIN, J., MAYBERRY, T., DONAHUE, C., FOPPE, L., BROWN, L., RIGGINS, C., & RYE, E. C., 2017. a Study of MAC Address Randomization in Mobile Devices and When it Fails. *Proceedings on Privacy Enhancing Technologies*, 365-383.

GU, X., WU, W., GU, X., LING, Z., YANG, M., & SONG, A., 2020. Probe Request Based Device Identification Attack and Defense. Sensors (Basel, Switzerland), 20(16), 4620. https://doi.org/10.3390/s20164620

IEEE, 2018. IEEE 802.11aq-2018 - Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Preassociation Discovery. Retrieved from IEEE Standard for Information technology - Telecommunications and information exchange between systems Local and metropolitan area networks https://standards.ieee.org/standard/802_11aq-2018.html

BRAVENEC, T., 2021. ESP32 Probe Sniffer. https://gitlab.com/tbravenec/esp32-probe-sniffer

LOH C. C., D., CHIA, Y. C., TAN C, P., & RI SENG, L., 2008. Identifying Unique Devices through Wireless Fingerprinting. Proceedings of the first ACM conference on Wireless network security (p. 10). Alexandria, VA, USA: Association for Computing Machinery.

BARBERA, M. V., EPASTO, A., MEI, A., KOSTA, S., PERTA, V. C., & STEFA, J., 2013, September. Dataset sapienza/probe-requests (v. 2013-09-10). Retrieved from CRAWDAD: https://crawdad.org/sapienza/probe-requests/20130910

BARBERA, M. V., EPASTO, A., MEI, A., KOSTA, S., PERTA, V. C., & STEFA, J., 2013. Signals from the Crowd: Uncovering Social Relationships through Smartphone Probes. Proceedings of the 2013 Conference on Internet Measurement Conference (p. 12). Barcelona, Spain: Association for Computing Machinery.

Please find this full paper in the following DOI

Sara Irina Fabrikant (2023) Neuroadaptive LBS: towards human-, context-, and task-adaptive mobile geographic information displays to support spatial learning for pedestrian navigation, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2258100

Exploratory analysis of mobile app usage in relation to distance from home

DONATELLA ZINGARO AND TUMASCH REICHENBACHER

Department of Geography • University of Zurich • Winterthurerstr. 190 • 8057 Zurich

Tel.: +49 (0)89 123 456 78 • E-Mail: donatella.zingaro@geo.uzh.ch

Keywords: mobile app usage, spatial dependency, tappigraphy, ambulatory assessment

Summary: *We present a first exploratory analysis of app usage collected from 38 participants with the tappigraphy approach. In addition to collecting tapping data of our participants, we registered the GPS locations during their phone sessions. Our analysis entails the density estimation of smartphone session usage and the inspection of potential effects of distance from the home location on participants' number of taps in apps, differences in the number of taps on map and other apps, and finally on time spent on map apps. We found different behavioural patterns of mobile app usage on an individual level. However, overall, there are no significant differences in tap density across map and other app categories over the distance from home. Nonetheless, we argue that these preliminary results are crucial to investigate app usage behaviour on smartphones further and put a solid basis on the validation of tappigraphy as a method in the field of LBS and GIScience.*

Introduction

Mobile devices, such as smartphones and tablets, have become pervasive in modern life. Many people use map apps as their primary source of geographic information for navigating, spatial decision-making, and problem-solving. Surprisingly, little is known about where, when, and how those map apps are used (Reichenbacher et al., 2022). Ecologically valid user studies on mobile map usage are rare, and map use is approximated by general and aggregated app download numbers or self-reports on smartphone usage. Both are not able to capture everyday map app usage in geographic space. Thus, knowledge of such behaviour is urgently needed in the fields of location-based services (LBS) and GIScience to support the user-centred design and development of future human and context-dependent map apps, as already addressed in several recent research agendas (see Huang et al., 2018, Thrash et al., 2019). To date, the most comprehensive field study on mobile map usage is the one from Savino et al. (2021), where data on Google Maps interactions were collected with a wrapper app *MapRecorder*. However, *MapRecorder* could not track in-app navigation and smartphone apps in general. Reichenbacher and colleagues (2022) recently used, for the first time, tappigraphy to obtain smartphone app usage data in everyday situations as a method for continuous, unobtrusive collection of 'natural', ecologically valid smartphone touch patterns. Tappigraphy has been developed in and, so far, used in the field of neuroscience (e.g., for quantifying hidden variables such as sleep, cognitive processing speed, and disease activity; Balerna & Ghosh, 2018; Borger et al., 2019; Duckrow et al., 2021; Huber & Ghosh, 2021). However, tappigraphy could be a promising ambulatory or ecological momentary assessment (EMA) method to explore mobile geographic information usage behaviour. Here, we used the same method as Reichenbacher et al. (2022) to collect and further analyse taps on apps in participants' smartphones to explore spatial and temporal tapping patterns and to detect behavioural patterns according to their distance from home location. With tappigraphy as EMA, we can now easily collect map app usage behaviour on an individual level. Analysing the distance-dependent extent and type of map app usage could provide insights into when people use map-based information and for which purpose (e.g., wayfinding, exploring, planning, etc.). Such knowledge could help in designing more customised map apps.

We begin our exploratory analysis by studying app usage in geographic space by exploring distance effects before investigating the differences between two main categories of apps (map apps and other apps) as the first step toward a better understanding of everyday map usage behaviour. We argue that this exploratory analysis is crucial to validating tappigraphy as a method to detect relevant information regarding mobile app usage behaviour and to set the basis for identifying typical usage scenarios and later clarifying generalisable interaction patterns, especially for mobile geographic information apps.

Methods

To shed some light on mobile app usage, we collected smartphone app usage data from 42 participants between February 2021 and April 2022. Participants were recruited through mailing lists, announcements through our community network and publishing the call on our department and university website. Participants did not receive any financial remuneration, but they participated in a raffle for a tablet. By registering on our project website and giving consent to the study terms, participants received a randomised code to be identified and were instructed to download the free MapOnTap app from the Google Play Store and install it on their smartphones. The app is only available for the Android operating system and is based on a tapping recording app called Tap Counter (QuantActions Sàrl) running in the background of their smartphones. Note that we have no knowledge about our participants than their assigned code by the ambulatory assessment approach of our data collection and by privacy protection. Participants were then using their smartphones as usual for at least two weeks so that we could record phone sessions. A unique ID phone session was generated each time, from when the phone was unlocked until locked again by the participant or the screen went to sleep mode. Data related to the taps on the active, foreground app includes the number of taps for each phone session, the start and the stop of the phone session, an app ID (i.e., the unique code identifying the active app and its category tapped on by the specific participant), the participant ID (i.e., random code attributed to each participant) and the device ID (i.e., random code of the device used by the participant). During our data collection campaign (February 2021 to April 2022), we recorded taps on participants' smartphones and GPS fixes as a series of timestamps occurring in an active phone session (i.e., when the tracking permission is on and permitted by the participant), accessing the smartphone's location sensors. GPS data was directly pushed to a table in our Postgres database on our server. Overall, we collected 52,688 phone sessions with an average duration of GPS recording (not active phone sessions but also when the phone is locked) of 54 minutes and a total of 2,978,096 timestamps with latitude and longitude coordinates with an average collection rate of 2.6 timestamps per minute. The tapping data collected by the MapOnTap app on participants' smartphones were pushed to the cloud platform operated by QuantActions every time a Wi-Fi connection was available. From the cloud platform, we downloaded the data and imported it to a Postgres database on a server in our secure University IT infrastructure. For tapping data, we have 119,713 phone sessions with an average duration of 5 minutes and 11 seconds and a total of 12,492,705 timestamps.
We imported the phone sessions with the tapping data and the phone sessions with geographic coordinates and timestamps from the Postgres database to two different pandas' data frames on Jupyter Labs. The Jupyter Labs environment is running on a different virtual server in our secure University IT infrastructure.
Our exploratory analysis investigates potential distance effects on smartphone usage behaviour. Thus, we first considered the geographic coordinates and the related timestamps only for each participant. Hence, we conceptualised and computed the home location for each participant as longitude and latitude mode of all GPS data from all participants' phone sessions. It seems a plausible approach as we assume that participants would spend most of their time at home. Next, we computed all distances from home to all GPS fixes (in kilometres) and categorised these distances into three main groups, close to home (~ 50 km), mid-range distance (~ 150 km) and far-range distance (> 150 km).
In the second part of the exploratory analysis, we included and combined information related to each participant's recorded taps and used apps. The apps are labelled according to Google Play Console[1]. Employing this categorisation, we found in total 32 distinct app categories

[1] https://support.google.com/googleplay/android-developer/answer/9859673?hl=en#zippy=%2Capps

used by participants (e.g., *Communication*, *Social*). After extracting tap and app data, we computed the number of all touches per session over the distance from home.
In the third part of our exploratory analysis, we also considered the map app category. We grouped all the app categories recorded into two main categories: the 'map app' category (i.e., apps categorised in the Google Play Console as *Travel and Local, Maps and Navigation*) and the 'other app' category (i.e., all the other apps categorised in the Google Play Console, such as *Communication*, *Game*, *Social*, etc.). Here, our tap data frame was split into map app categories, with a total of 489231 taps and other app categories with a total of 12,003,474 taps. We computed and plotted the KDE of touches on map apps and other apps over distance from home. Next, we calculated and plotted the time spent on map apps over the distance from home for each participant. Time spent on map apps was calculated for each participant's phone session taking the start and end of each phone session and then subdividing the phone session into subsections containing the start and the stop of map taps interactions. The resulting time was computed by aggregating the time difference between beginning and the end map tap.
Finally, we computed for each participant the densities of map app taps and other app taps normalised by the total recording period time within the convex hull spawn by the GPS fixes. We assume that the GPS fixes registered by the participants' smartphones define the aggregate activity spaces. The densities of taps within these activity spaces should allow us to detect deviations from the expected linear growth of the number of taps with larger activity spaces.

Results and Discussion

From the initial pool of 42 participants, we excluded one participant with empty coordinate entries from the GPS data frame. Next, three additional participants were excluded because they had empty entries in the tapping data frame. Our final sample consisted of 38 participants.

The median of participants' median distance from home is 8.84 km, and the median number of taps is 24. We further computed the median ratio of taps over distance from home for each participant (MIN: .016; MAX: 31875; M: 1287; SD; 5281). The median of this median ratio is 2.63 taps per km. Successive, we split the 38 participants into two groups above and below the median taps over distance ratio. Since our data are not normally distributed, we applied a non-parametric Wilcoxon test to check for distance dependency of the tapping ratio. The test revealed that the central tendency for the two groups is significantly different ($Z = 0$, $p < .001$), which means that the tapping ratio depends on the distance from home.

Thus, to better investigate smartphone app usage behaviour, we included in the following step the recorded tapping and app data of participants in conjunction with the GPS data. We plotted the number of taps per individual phone session over the computed distances from the estimated home. As such, we could better identify all our participants' different app usage patterns over spatial distance. Figures 1–3 depict the number of taps per phone session for three selected participants that show distinct patterns of tapping depending on the distance ranges from their respective home locations. These plots allow us to inspect distance effects and distinct app usage patterns visually.

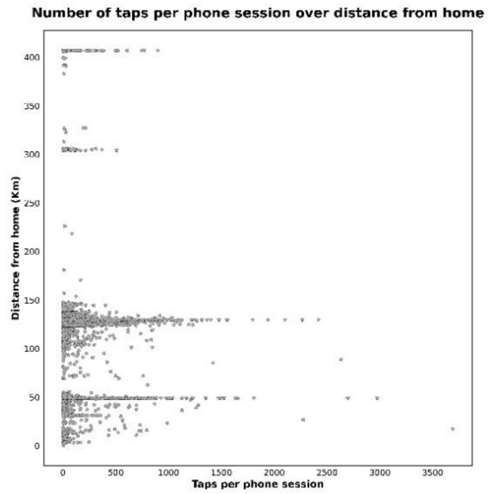


Fig. 1: Number of taps per phone session over distance (in kilometres) from estimated home location of P1.

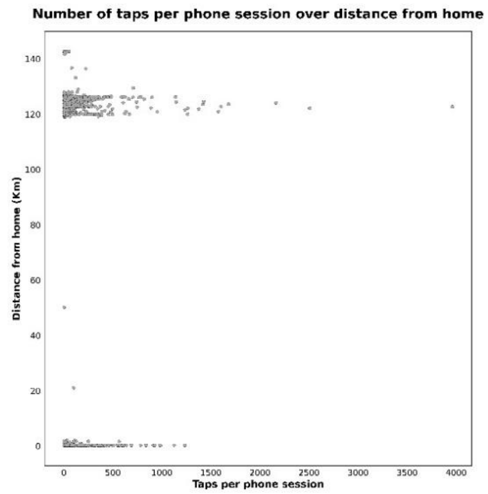


Fig. 2: Number of taps per phone session over distance (in kilometres) from estimated home location of P2.

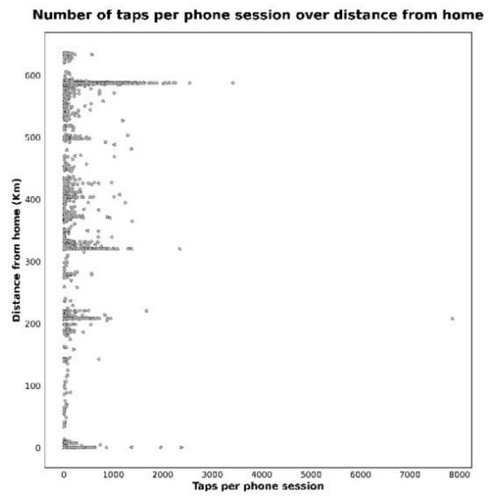


Fig. 3: Number of taps per phone session over distance (in kilometres) from estimated home location of P3.

Two clear patterns manifest most frequently across all participants. The first pattern we identified is when smartphone apps are mainly used at home, close (~ 50 km), and in a mid-distance range from home (~150 km) (Fig. 1 and 2). This pattern is shaped as a bimodal distribution. The first bump shows that the highest number of phone sessions are recorded near or around the home location, and the second bump is around the mid-distance range. Interestingly, the number of taps does not increase with distance from home. Instead, the numbers of taps stay low (~100 taps), suggesting that participants do not use their phones continuously but prefer a short average session when accessing an app. This result aligns with findings on tappigraphy reported in (Reichenbacher et al., 2022).

P1 and P2 were not the only two cases of bimodal distribution of taps, which leads us to think that individuals might have a peculiar behaviour when tapping and using their smartphones over space. Further analysis should consider the taps over distance from home for our sample to detect a significant behavioural pattern.

The second pattern identified in the exploratory analysis shows a distribution of phone sessions all over the distance range (Fig. 3). This might reflect an extensive usage behaviour of the smartphone and its apps regardless of spatial factors, such as proximity to the home location. Moreover, Figure 3 shows an increase in app taps when they happen further away from home. This pattern reflects one of our crucial hypotheses: increasing taps far from home could be associated with increased taps on map apps. To investigate our assumption, we divided the app categories registered into two main categories map apps and other apps. For the analysis, we have also considered that the number of taps on map apps is much smaller than taps on other apps (see Do et al., 2011, Carrascal & Church, 2015; Fonseca et al., 2021; Reichenbacher et al., 2022). We plotted the KDE of taps on the two distinct app categories over the identified distances from home. Surprisingly, we have found that the increased number of taps with growing distance from home is related to map app usage for some participants (see Fig. 4–6). The additional planned analysis will unravel a finer-grained categorisation of the apps to get more insights into which apps are used where and when. Such knowledge about the usage of apps either near home or, on the opposite, far from home will allow us to identify typical app usage scenarios across space for individuals.

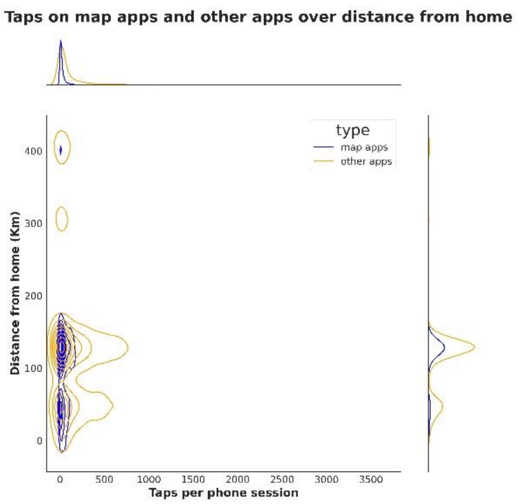


Fig. 4: Kernel Density Estimation (KDE) of number of taps per phone session on map apps and other apps over distance from home for P1.

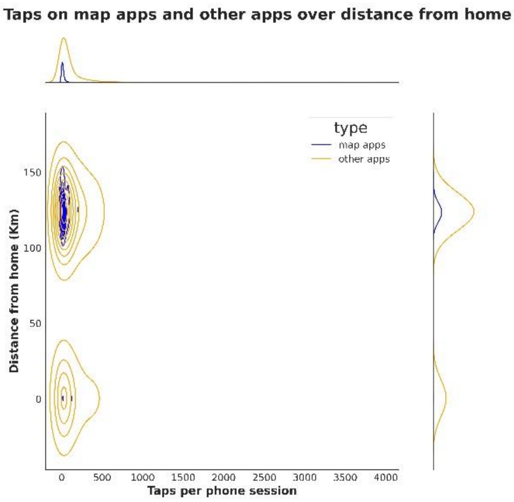


Fig. 5: Kernel Density Estimation (KDE) of number of taps per phone session on map apps and other apps over distance from home for P2.

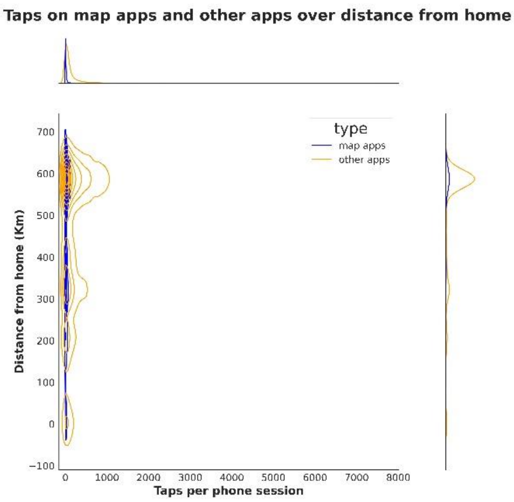


Fig. 6: Kernel Density Estimation (KDE) of number of taps per phone session on map apps and other apps over distance from home for P3.

Lastly, we wanted to focus on the number of taps occurring across the spatial distance from home and the time spent on map apps and other apps dependent on distance from home. This will help us clarify if map apps are used for a longer time when far from home to navigate unfamiliar environments. Time spent on app categories over the distance from home was calculated for all participants for each phone session and then plotted over the distance from home. Figures 7–9 show the results for the three selected participants. The left panel shows the time spent on map apps, and the right panel displays the time used for other app categories. These patterns reveal different tendencies encountered across participants when exploring our data. However, the tapping patterns and time usage are not substantially different for the two main categories (map and other apps).

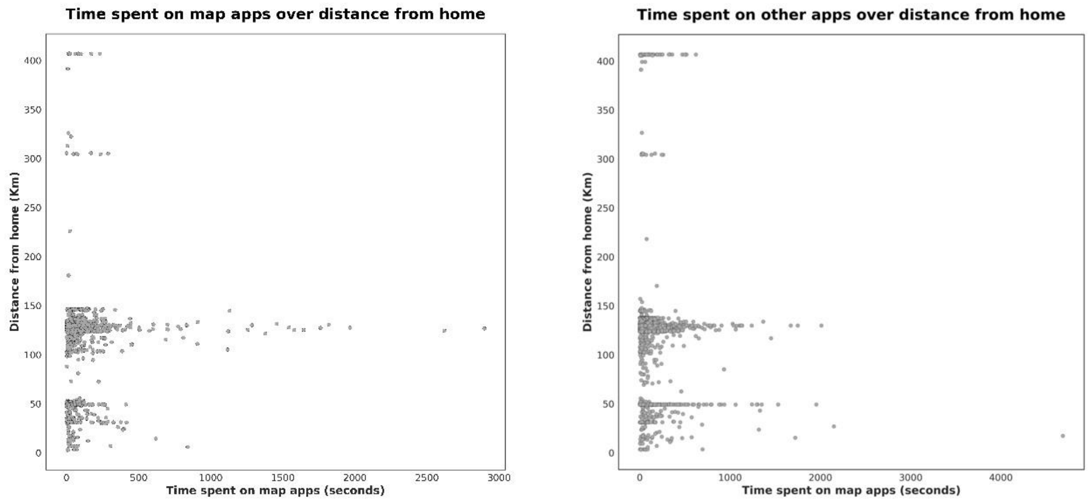


Fig. 7: Scatter plot of time spent (in seconds) over distance from home (in kilometres) on map apps (left panel) and other apps (right panel) for P1.

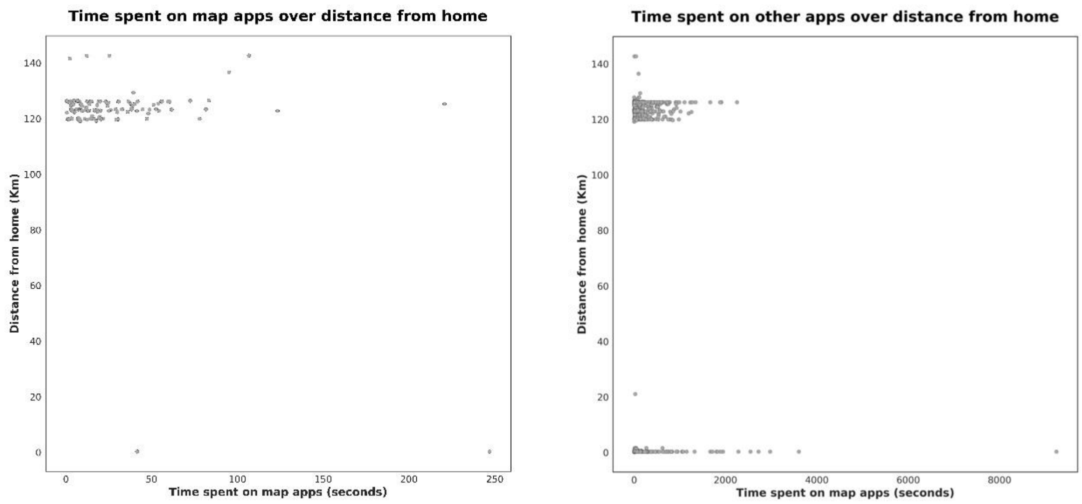


Fig. 8: Scatter plot of time spent (in seconds) over distance from home (in kilometres) on map apps (left panel) and other apps (right panel) for P2.

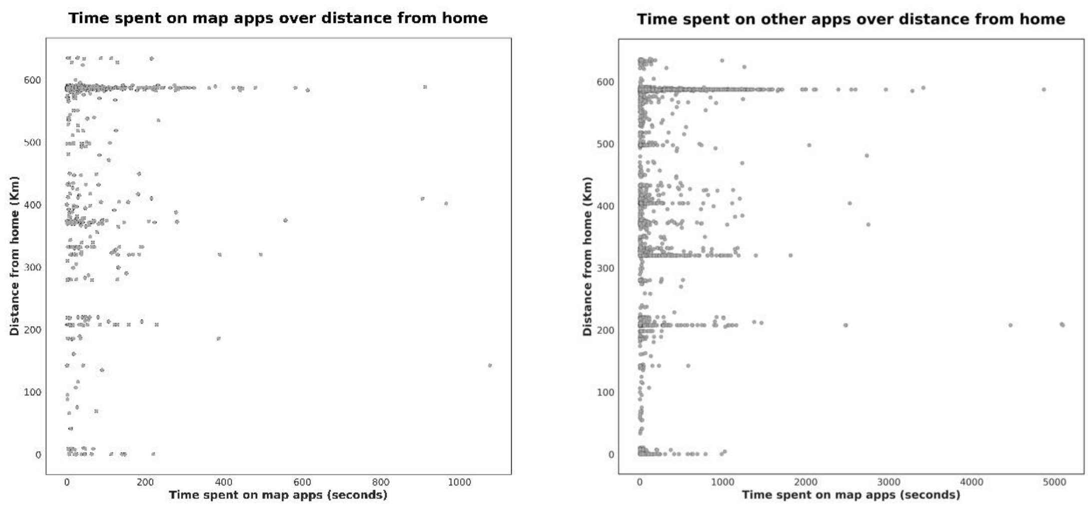


Fig. 9: Scatter plot of time spent (in seconds) over distance from home (in kilometres) on map apps (left panel) and other apps (right panel) for P3.

Figure 7 shows longer map usage time when in mid-range distance from home, suggesting various usage modes, such as planning, checking own position, or navigation and route following. Further analysis should consider different time granularities (e.g., daytime, weekdays, etc.) to reveal more distinct patterns and check whether they align with previous literature analysing app usage information from users of Android-powered mobile devices (e.g., Böhmer et al., 2011). In figure 8, we can see that P2 is spending time on maps only at a mid-range distance. This makes us confident of finding in future analysis stable patterns of mobile map app usage when further away from home in what is likely to be more unfamiliar environments. Figure 9 supports those analyses by showing a more extended time spent on map apps over a greater distance from home. Again, including different time granularities and a larger sample in our analysis could help us better to understand the purpose of mobile map app usage.
To verify our hypothesis that densities of taps within activity spaces deviate from an expected linear growth of the number of taps with larger activity spaces, we first split the participants into two groups: above and below the median distance from home for both map taps and other app taps. Then, after verifying our sample's non-normal distribution, we performed the Wilcoxon test on the density of map taps: $Z = 17.0$; $p < .001$. Similarly, we computed Wilcoxon test on density of other app taps: $Z = 15.0$; $p < .001$. Both results showed a significant difference between the densities of taps of the two groups, implying a high dependency on distance from home. Thus, because both main categories of apps (i.e., map and other) showed similar significant results when performing the Wilcoxon test, we can conclude that there is no significant difference across these two categories and hence for the density of taps over the home distance. These results are consistent with the patterns visible in Fig. 7–9.

Conclusions

With our exploratory analysis of smartphone app usage, taking taps on the smartphone touch screen as a proxy for app usage, we could show some frequently occurring behaviour patterns across participants. The tappigraphy method revealed distinct mobile usage patterns in phone sessions when app taps were plotted over distance from the estimated home locations. Furthermore, when taps on apps were divided into two main categories (map apps and all other apps), we could better identify the spatial dependency of map app usage behaviour. These results align with our hypothesis that map apps are used more with increased distance from home. However, when comparing the densities of taps for the map and other apps, we could not see any significant change in the tapping patterns. This suggests that while moving further from the home location, apps are used independently from the categories in which they belong. However, finer-grained categorisation of apps could reveal usage patterns that we failed to identify here. This work only presented results from an exploratory analysis using tappigraphy methodology. For a future study, we will expand our analysis in different directions. To reveal more distinct spatiotemporal map usage patterns, we will include a finer-grained categorisation of the apps and different time granularities. In conclusion, coherently with our current findings and the initial state of this exploratory analysis, we argue that tappigraphy could be considered a reliable EMA method for understanding mobile geographic information usage behaviour and eventually support the user-centred design and development of future human and context-dependent map apps.

Acknowledgement

We thank QuantActions Sàrl for their support in our data collection campaign and Enea Ceolini for the design, development, and implementation of the MapOnTap app that allows us to obtain location data of the smartphone app usage.

References

BALERNA, M., & GHOSH, A., 2018: The details of past actions on a smartphone touchscreen are reflected by intrinsic sensorimotor dynamics. Npj Digital Medicine, 1, 4. https://doi.org/10.1038/s41746-017-0011-3

BORGER, J. N., HUBER, R., & GHOSH, A., 2019: Capturing sleep–wake cycles by using day-to-day smartphone touchscreen interactions. Npj Digital Medicine, 2(1), 73. https://doi.org/10.1038/s41746-019-0147-4

BÖHMER, M., HECHT, B., SCHÖNING, J., KRÜGER, A., & BAUER, G., 2011: Falling asleep with Angry Birds, Facebook and Kindle: A large scale study on mobile application usage. https://doi.org/10.1145/2037373.2037383

CARRASCAL, J. P., & CHURCH, K., 2015: An in-situ study of mobile app & mobile search interactions. https://doi.org/10.1145/2702123.2702486

DUCKROW, R. B., CEOLINI, E., ZAVERI, H. P., BROOKS, C., & GHOSH, A., 2021. Artificial neural network trained on smartphone behavior can trace epileptiform activity in epilepsy. IScience, 24(6), 102538. https://doi.org/10.1016/j.isci.2021.102538

DO, T. M. T., BLOM, J., & GATICA-PEREZ, D., 2011: Smartphone usage in the wild: A large-scale analysis of applications and context. https://doi.org/10.1145/2070481.2070550

FONSECA, F., CONTICELLI, E., PAPAGEORGIOU, G., RIBEIRO, P., JABBARI, M., TONDELLI, S., & RAMOS, R., 2021: Use and Perceptions of Pedestrian Navigation Apps: Findings from Bologna and Porto. ISPRS International Journal of Geo-Information, 10(7), 446. https://doi.org/10.3390/ijgi10070446

GHOSH, A., PFISTER, J.-P., & COOK, M., 2017: Optimised information gathering in smartphone users. ArXiv:1701.02796 [Physics.Soc-Ph].

HUANG, H., GARTNER, G., KRISP, J., RAUBAL, M., & VAN DE WEGHE, N., 2018: Location based services: Ongoing evolution and research agenda. Journal of Location Based Services, 12(2), 63–93. https://doi.org/10.1080/17489725.2018.1508763

HUBER, R., & GHOSH, A., 2021: Large cognitive fluctuations surrounding sleep in daily living. IScience, 24(3), 102159. https://doi.org/10.1016/j.isci.2021.102159

REICHENBACHER, T., ALIAKBARIAN, M., GHOSH, A., FABRIKANT, S., 2022: Tappigraphy: continuous ambulatory assessment and analysis of in-situ map app use behaviour. Journal of Location-Based Services. http://dx.doi.org/10.1080/17489725.2022.2105410.

RIEGELSBERGER, J., & NAKHIMOVSKY, Y., 2008: Seeing the bigger picture: A multi-method field trial of google maps for mobile. https://doi.org/10.1145/1358628.1358655

SAVINO, G.-L., STURDEE, M., RUNDÉ, S., LOHMEIER, C., HECHT, B., PRANDI, C., NUNES, N. J., & SCHÖNING, J., 2021: MapRecorder: Analysing real-world usage of mobile map applications. Behaviour & Information Technology, 40(7), 646–662. https://doi.org/10.1080/0144929X.2020.1714733

THRASH, T., LANINI-MAGGI, S., FABRIKANT, S. I., BERTEL, S., BRÜGGER, A., CREDÉ, S., DO, C. T., GARTNER, G., HUANG, H., MÜNZER, S., & RICHTER, K.-F., 2019: The Future of Geographic Information Displays from GIScience, Cartographic, and Cognitive Science Perspectives (Vision Paper). 142, 19:1-19:11. https://doi.org/10.4230/LIPIcs.COSIT.2019.19

Replication of Wayfinding Studies in Different Geographic Areas. A Simulation Study

BARTOSZ MAZURKIEWICZ AND IOANNIS GIANNOPOULOS

Research Division Geoinformation • TU Wien • Gusshausstraße 27-29• 1040 Vienna, Austria

E-Mail{bartosz.mazurkiewicz, ioannis.giannopoulos}@tuwien.ac.at

Keywords: Route Selection, Human Wayfinding, Experiments, Replicability

Summary: *Replication of real-world wayfinding studies is not a trivial task. Even less if it is to be replicated in a different geographic environment. The selection of one or several routes is one of many decisions to be made. Only recently (2021), a reproducible, systematic and score-based approach for route selection for wayfinding experiments was published. Besides allowing for selecting a route within a selected experimental area, it claims to be able to find similar routes in different geographic areas. However, it remains unclear if similar, according to this route selection framework, routes lead to similar study results. In order to answer this question, an agent-based simulation comparing Turn-by-Turn and Free Choice Navigation approaches (between-subject design) is run in one European (Vienna) and one African (Djibouti City) city. First, a route in Vienna is selected and, second, the 5 most and the 5 least similar routes in Djibouti City are found. These routes are used in the simulation in order to scrutinize if more similar routes lead to more similar results regarding the arrival rate as a metric. The results suggest that the route selection framework is suitable for replication studies for the Turn-By-Turn navigation approach but needs further improvement for the Free Choice Navigation approach by adding features describing the neighborhood of the route.*

1. Introduction

The replication of studies is not a trivial task, as many factors need to be considered and kept as similar as possible to make the results comparable. The route selection is crucial for replicating wayfinding studies. There are two possibilities regarding the experimental area. It can be kept constant, although some elements of the environment may have changed over time and potentially impact study results. The second option is to replicate a wayfinding study in another geographic area. In the second case, the route selection task is not as simple as in the first case (using the same route). The routes from both studies, the original and the replicating one should be similar regarding the wayfinding task.

We recently presented (2021) a framework [14] that allows systematic route selection, i.e., how to select a route from a given experimental area with many potential routes. Furthermore, we hypothesized that this framework would increase the replicability of wayfinding studies by finding similar routes in different geographic areas. If this assumption can be verified, then the above-mentioned problem of selecting similar routes in different geographical areas can be solved or at least mitigated. Therefore, we will use the previously proposed route selection framework, first, to identify an average-based [14] route in a European city (Vienna) and, second, to find the most and the least similar routes in Djibouti City in Africa. Two navigation systems (see Section 3.2) will be compared on these routes with respect to the arrival rate. Since the framework can capture route characteristics, more similar routes in Djibouti City should lead to more similar results to those achieved in Vienna. As in our previous study [13], this hypothesis will be scrutinized through a simulation study.

The contribution of this work is two-fold: First, the suitability of the route selection framework for replication studies is investigated. Our results suggest the ability of the route selection framework to support replication studies in other geographic regions. Furthermore, it should increase the comparability of wayfinding studies if the selected experimental areas with their respective routes are similar enough. Second, we shed light on the importance of route selection in wayfinding studies by analyzing the arrival rates on single routes.

2. Related Work

In this section, we discuss relevant literature, first, about reproducibility in the domain of GIScience in general and, second, about replication in the wayfinding domain. In this work, the terms reproducibility and replication are used in the sense of Claerbout/Donoho/Peng [2]. Reproduction means recreating the results with the same methods and input data that the authors provide. The related concept of replication means coming to the same conclusion by conducting a new study.

2.1 Reproducibility in GIScience

Reproducibility has seen considerable interest in the GIScience domain within the last years (e.g., [9, 3]). Ostermann and colleagues assessed 87 papers from GIScience conferences between 2012 and 2018 regarding reproducibility [17]. None of the assessed works was easily reproducible. This study replicated a study considering the AGILE conference [16]. In conclusion, both conference series are similar regarding reproducibility. Konkol and colleagues conducted a study about computational reproducibility in geographic research [10]. They studied the understanding of open reproducible research (ORR) through surveys, interviews and a focus group. They found that the meaning of ORR diverges considerably among the participants of the European Geosciences Union General Assembly 2016. Furthermore, the authors tried to reproduce the results and figures of 41 open access articles from Copernicus and the Journal of Statistical Software. They encountered technical issues of different severity levels in 39 works.

2.2 Replication in Different Geographic Areas in the Wayfinding Domain

Several studies have been conducted replicating real-world studies in virtual environments. Kuliga and colleagues [11] conducted a wayfinding study in a building and then replicated it three times in different virtual replicas. All four conditions yielded similar results regarding superfluous distances and absolute angular pointing errors. Savino and colleagues compared wayfinding in real-world and virtual environments [20]. They found differences between both navigation aids (paper map and smartphone) in both conditions regarding stopping time and task load, among others. No new route was selected in both studies, as the virtual environment reflected the real world.

Wayfinding studies replicated or conducted in a different geographic area are usually based on questionnaires rather than actual wayfinding studies (see e.g., [12, 15]). To the best of our knowledge, there is no work replicating an actual pedestrian wayfinding task in a different geographic area. One reason for this might be the difficulty of selecting appropriate routes. Our work contributes to the realization of replication studies in the wayfinding domain, which are conducted in different geographic areas by facilitating the route choice.

In many wayfinding experiments (e.g., [6, 5]) in which at least two navigation systems are compared, one of the conditions is a map-based Turn-By-Turn navigation approach (e.g., Google Maps). The replication of this widespread baseline condition is rather simple (App availability) but still time-consuming. Given that many empirical results are available for this and other approaches, there might be a possibility to avoid the replication of baseline approaches in every experiment. This would allow comparing novel systems against existing ones by reproducing the experimental setup but having to collect the results for the novel approach only.

3. Experimental Setup

In this section, the agent-based simulation study with its two navigation systems, Turn-by-Turn (TBT) and Free Choice Navigation (FCN), is described in detail. We will elaborate on both experimental areas and all potential routes with pre-defined features. As in our previous work [13], the study follows a between-subject design with 6000 agents. The choice between a between-subject or within-subject design is of less importance, as long as both groups do not differ significantly regarding their environmental spatial abilities (see Section 4), which mainly influence the performance (see Section 3.2).

3.1 Experimental Areas

As the original experimental area (source city), the city center (surface area 2.5 km^2) of Vienna is chosen (see Figure 1). According to the classification by Thompson et al. [22], the network layout is of type high transit. The city for which suitable routes for a replication study need to be found is Djibouti City in Africa (see Figure 2), which is of network type irregular [22]. The selected experiment area is of similar size (surface area 2.27 km^2) and lies in the western Part of Djibouti City (see Figure 2). The size of the experimental areas is of less importance, as long as there are routes of the desired length (see Section 3.3). Bigger experimental areas mean more potential routes and result in longer computation times.

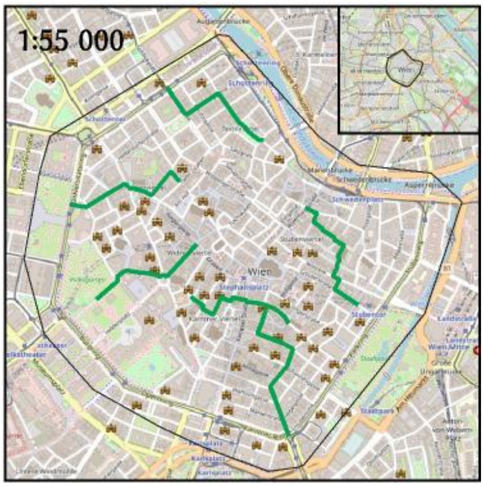


Figure 1: The experimental area in Vienna with six sample routes. Basemap © OpenStreetMap.

Figure 2: The experimental area in Djibouti City with six sample routes. Basemap © OpenStreetMap.

For both experimental areas, the raw network data were downloaded from OpenStreetMap (OSM)[1]. The intersections and their characteristics were calculated using the Intersections Framework [4], whereas street segments between two intersections were extracted with a custom script. For both areas, a networkx graph was created, which was used for the simulation.

3.2 Navigation Systems

Following our previous work [13], we compare the same two navigation approaches, namely

[1] https://www.openstreetmap.org, last access March 25th, 2022

Free Choice Navigation (FCN) and Turn-by-Turn (TBT). The primary reason to use a navigation assistance system is the desire to reach a defined destination. Therefore, the arrival rate is chosen as the success metric. An agent successfully reaches its destination if the walked distance does not exceed 150% of the shortest path length [13].

3.2.1 Turn-by-Turn (TBT)

In this condition, the agent will be guided along the shortest path between origin and destination and receives only at turning points navigation instructions. It is a popular approach that is often used as baseline in navigation experiments (e.g., [8, 21]). Whenever agents have to go straight ahead (continuation within a 20° cone concerning the current walking direction) at a junction, then no instruction is issued, and the agent will not turn. Every agent has a fixed probability to interpret generic navigation instructions correctly, which ranges between 0.8 and 1. We expect such a high probability [13] because navigation instructions are followed every day by millions of users. The agent interprets a turning instruction using a weighted random choice: The branch indicated in the instruction obtains a weight equal to the agent's probability to interpret generic navigation instructions correctly. The remaining probability is distributed equally over all remaining branches, excluding the one indicated in the instruction and the most recently taken branch. Once the agent reaches the destination, the trial ends.

3.2.2 Free Choice Navigation (FCN)

Free Choice Navigation is a navigation paradigm aiming for more freedom of choice during navigation, trying to balance the number of free choices, given instructions and a maximum allowed route length [13]. The following example shows the working mechanism: Anna, a good wayfinder, navigates to an art gallery. Before the navigation starts, Anna receives information about the beeline direction and distance to the art gallery. The system does not issue any instructions at the first two junctions because the beeline direction should still be clear to the user after such a short period. In this situation, Anna decides on her own which branch to take. The third junction, however, is rather complex and has six branches. Anna is quite sure about the beeline direction towards the art gallery, but two branches seem to be good choices to her. Based on internal computations which take her spatial abilities and the environmental structure into account, the navigation system becomes aware of this difficulty (see our previous work [13]). Consequently, Anna receives an instruction because one of the branches results in a considerable deviation from the acceptable route length. The instruction is interpreted correctly and Anna continues her way to the art gallery.

This example illustrates which components influence the internal computations of the navigation system: the user's environmental spatial abilities, the features of the current junction and the already traversed route. If an instruction is issued, a similar procedure as above applies, with the difference that the last taken branch is not excluded but has a lower probability of being taken. Another difference is that the agent's probability to interpret the generic navigation instruction correctly (as well between 0.8 and 1) depends linearly on its environmental spatial abilities. For more details, please refer to the original paper [13].

3.3 Route Selection

In this section, an average-based route in the source city and the most/least similar routes in the target city are selected. Our previously proposed route selection framework was used for these tasks [14].

As pre-emptive criteria [14], we set the route length between 550 m and 1000 m (see e.g., [18, 19]) and the number of decision points on a route to 12 (according to OSM) to avoid trivial route length. Only shortest paths were considered suitable for our experimental design. Given that the two navigation approaches depend on the geometry of the route and the network (see Section 3.2), geometry-based routes features were selected [14]: average number of branches, number of n-way intersections (e.g., 3-way intersections), regularity of decision points [4], number of right, left and non-turns and length-related features (average, median and standard deviation of segment lengths and total route length). All features were equally weighted. To find all possible routes meeting the set criteria, we followed the original paper

[14] and used SageMath 9.1 with its SubgraphSearch function[2]. In Vienna, 11737 shortest paths meeting the above-mentioned criteria were found and 9064 in the experimental area in Djibouti City.

3.3.1 Vienna

For every route in Vienna, the weighted Euclidean distance (called score) to the hypothetical route, which shows closest to average values for all criteria, was calculated [14]. Four routes yielded a minimal score of 0.12 (0 would indicate a perfect match). Actually, there are only two distinct routes, since every route is present twice. Two distinct routes traversed from start to destination and vice versa result in four routes. All four routes are very similar, and they differ regarding the direction and a turn while entering a square (see Figure 3). Due to these similarities, no route could be defined as better than the others, and consequently, all four routes are considered suitable.

For each of these routes, the five most and five least similar routes in Djibouti City were found using the framework. Five routes were chosen due to two reasons. First, arrival rates for five routes are more representative than considering one route only. Second, five seems a reasonable number in the route selection process because higher-ranked routes may not always be suitable for the experiment due to uncaptured characteristics in the route features (e.g., data not available). In this case, lower-ranked routes need to be considered too. The route selection framework is an assistance system, and local knowledge will always help to make the final decision, potentially excluding higher-ranked routes. This expert knowledge does not impede reproducibility, if the decision is well documented.

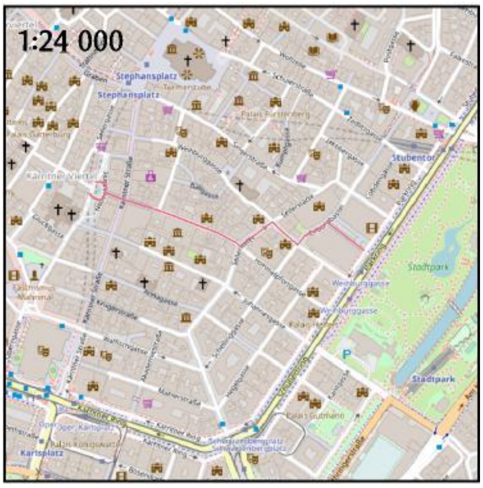


Figure 3: Routes in Vienna. The four routes differ in direction and a turn while entering a square. Basemap © OpenStreetMap.

3.3.2 Djibouti City

While searching for the most and least similar routes in Djibouti City, two further features were added to increase the similarity to the source routes. Both features concentrate on the order of one of the above-mentioned features (see Section 3.3). The sequence of right, left and non-turns (e.g., 'rnrlrnl') and the sequence of the cardinality of decision points (e.g., '3334343') along the route were considered, as they potentially influence the simulation results (e.g., more branches lead to more difficult decisions).

In Vienna, the Euclidean distance was calculated between every route and a hypothetical average route (hence the term average-based). In Djibouti City, the latter is substituted by the routes found in Vienna, respectively (see Section 3.3.1). As the two newly added features are strings, the Levenshtein distance was used to calculate the difference.

[2] https://doc.sagemath.org/html/en/reference/graphs/sage/graphs/generic_graph_pyx.html, last access March 4th, 2022

For each of the four considered routes coming from the source city, the five most and least similar routes in the experimental area in Djibouti City were calculated. The Euclidean distances for the most (M=0.903, SD=0.096, MIN=0.683, MAX=1.022) and least (M=3.334, SD=0.417, MIN=2.49, MAX=3.641) similar routes differ considerably.

4. Simulation Results

For each route, the whole simulation was run 100 times in order to counterbalance the influence of the weighted random choice function (see Section 3.2). Each route was walked by two (TBT and FCN) groups of 3000 agents. The presented numbers are the means of the corresponding route(s) for all 100 runs (different seeds). To ensure that the common ability of agents to interpret navigation instructions correctly did not influence the results, a Wilcoxon Signed-Rank Test on these abilities of the agents was performed. No significant (α = .05) differences between both conditions were found n = 3000 (Z = .00, p = .99, r = .00). The general influence of the these abilities on the Free Choice Navigation approach was discussed in our previous paper [13]. For each city, the parametrization (FCN) with the best balance between arrival rate and freedom of choice was used [13].

Vienna			Djibouti City			
			Most Similar Routes		Least Similar Routes	
Route	TBT	FCN	Mean TBT	Mean FCN	Mean TBT	Mean FCN
0	0.962	0.954	0.923	0.857	0.854	0.916
1	0.966	0.96	0.953	0.905	0.846	0.909
2	0.967	0.932	0.962	0.906	0.856	0.914
3	0.951	0.953	0.956	0.909	0.854	0.916

Table 1: Arrival rates for four equivalent (Euclidean distance score) routes in Vienna and their five most/least similar counterparts in Djibouti City. TBT - Turn-By-Turn, FCN - Free Choice Navigation, Mean - mean for 5 routes. The figures are rounded to three decimals.

4.1 Vienna

In the European city, both navigation systems reached a high arrival rate of around 0.95 (see Table 1). On three routes (0-2), TBT led more agents to the respective destination than FCN. On one route (3), FCN performed better than TBT. In general, the achieved arrival rates in Vienna are very similar for both navigation systems.

4.2 Djibouti City - Turn-By-Turn

For agents using the TBT navigation system, the most similar routes in Djibouti showed an arrival rate of around .95, which is close to the arrival rate in Vienna (see Table 1). The first route (0), however, is an exception, having a lower arrival rate of .923. The least similar routes in Djibouti showed an arrival rate of around .85, representing a considerable difference to both the most similar routes and the routes from the source city. For every route from the source city, the most similar routes in the target city yielded more similar results than the least similar routes.

4.3 Djibouti City - Free Choice Navigation

For agents using the FCN navigation system, the most similar routes in Djibouti showed an arrival rate of around .9, which is different from the arrival rate in Vienna (around 0.95, see Table 1). The first route (0), again, is an exception having a lower arrival rate of .857. The least similar routes showed an arrival rate of around 0.91, similar to the most similar routes. Moreover, the least similar routes in Djibouti yielded higher arrival rates than the most similar routes.

4.4 Djibouti City - TBT versus FCN

Comparing both navigation systems on the five most similar routes in Djibouti City shows that more agents reached their destination with TBT than with FCN. The opposite is observed while considering the least similar routes. In this case, FCN is superior to TBT regarding the arrival rate (see Table 1).

5. Discussion and Limitations

This section will discuss the results by comparing the arrival rates between and within cities, navigation approaches and the most and least similar routes. Furthermore, we discuss the limitations of our work.

The four selected routes in Vienna yielded similar arrival rates for both navigation systems (see Table 1). Only one route (2) led to a bigger difference of around 3%. This is not in line with the original work [13] in which TBT had, on average, a 5% higher arrival rate (100% vs. 95%). This indicates that route selection is crucial in experimental design because it can change the drawn conclusions and the outcome of a wayfinding study. For the TBT condition in Djibouti City, the route selection framework helped to find routes that yield, on average, a similar arrival rate as the corresponding source route. The least similar routes yielded considerably worse results (around 85%) compared to both the source routes in Vienna and the most similar routes in Djibouti City. This indicates the suitability of the route selection framework with the selected route features, as the lower-ranked routes yielded less similar results than higher-ranked routes. As Vienna and Djibouti City represent quite different layout types [22], we expect the framework to work as well in other geographic areas.

The FCN condition in Djibouti City shows a different picture, in which both the most similar and the least similar routes yielded high arrival rates but not as high as the source routes (see Table 1). Moreover, the least similar routes yielded better results in terms of arrival rate than the most similar routes. This can be explained by the interplay between the chosen route features and the navigation approach. One of the ideas of Free Choice Navigation is to give more freedom to the wayfinder. This increases the chances of not taking the shortest path, which is supposed to be taken in the TBT approach. The simulation data support this hypothesis (see Table 2).

Vienna			Djibouti City			
			Most Similar Routes		Least Similar Routes	
Route	TBT	FCN	Mean TBT	Mean FCN	Mean TBT	Mean FCN
0	107	597	79	344	52	282
1	105	601	85	658	58	242
2	121	380	89	375	57	190
3	139	398	67	329	52	282

Table 2: Number of uniquely walked routes taken by successful agents for four equivalent (regarding the Euclidean distance score) routes in Vienna and their five most/least similar counterparts in Djibouti City. TBT - Turn-By-Turn, FCN - Free Choice Navigation, Mean - mean for 5 routes. The figures are rounded to integers.

In the FCN condition, more unique routes are taken by successful agents in both Vienna and Djibouti City. With an increasing number of unique routes, the neighborhood around the route plays a more vital role. A route might be easy to navigate, but once a navigation error occurs, the wayfinder might find itself in a difficult to navigate area due to complex junctions, dead-ends or detours [1]. The selected properties (see Section 3.3), however, regard route properties only, without considering the neighborhood of the route itself. The route selection framework could be improved by including additional features, which capture the previously used characteristics but adapted for the neighborhood. Completely new features like centrality measures (graph theory) calculated for the route neighborhood could also help to improve the process of finding similar routes. This could be as well a first step to tackle the problem of conducting the baseline condition over and over again in wayfinding experiments (see Section 2.2). Previously collected empirical data could be used as a proxy if the neighborhoods and routes are highly similar.

However, the definition of such a neighborhood is not a trivial task and depends on the navigation system. Some routes are more likely to be taken with a given navigation system. We suggest incorporating features describing this neighborhood while considering the navigation system to define its spatial extent. One possibility to define the spatial extent of the route's neighborhood is the Potential Route Area (PRA) [8]. However, the PRA is based on shortest paths only, which are not necessarily taken.

The selected metric is important too. Regarding the number of unique routes (see Table 2), the results are as expected, more similar routes yielded more similar results than less similar routes. Regarding the arrival rate, the results are partially in line with our expectations (see Table 1). Therefore, the selected route features should consider the navigation system and the success metric.

The achieved arrival rates in Djibouti City are not entirely in line with the previously conducted simulation study [13]. Our study used 40 (Djibouti City) routes instead of the whole route population as in the original paper. A wayfinding study is usually conducted with a small-sized subsample of routes. The differences within the cities (see Table 1) and between our study and the original work [13] suggest that the selected route can impact study results (see Section 6).

5.1 Limitations

We could have added more complexity to the simulation with respect to the original study, but we wanted to keep our results comparable. In order to find similar routes, other similarity metrics could have been used. Toohey and Duckham [23] compare four different trajectory similarity measures, but all of them rely purely on route geometry. Han and colleagues used deep learning to calculate route similarity [7]. The authors, however, define the similarity based on node-wise distance over the underlying spatial network, although their architecture incorporates information about direct neighbors for a node, whose importance can be set by a parameter. In contrast to the selected route selection framework [14], however, the resulting similarity is not readily explainable.

6. Conclusion and Future Work

In our work, we wanted to verify if the proposed route selection framework can find similar routes in different geographic areas and, thus, make it suitable for replication studies. Our results reveal the suitability for the widespread Turn-By-Turn navigation approach and suggest the incorporation of further neighborhood features into the framework in order to work with navigation approaches that cover more potential routes between start and destination like Free Choice Navigation. This work is a first step towards the replication of wayfinding studies in different geographic areas.

For future work, there are several strands to follow. Further success metrics needs to be tested with our approach to see whether the results are applicable beyond the arrival rate and the number of uniquely walked routes. The definition of the neighborhood for a route is an open problem. We believe that it should depend on the tested navigation system. Furthermore, features describing this neighborhood are to be defined and verified. Our results suggested the importance of route selection on study results. We will scrutinize this hypothesis with a further simulation study in which we will run a wayfinding experiment on all suitable potential routes within the experimental area and compare the results. A further research direction is the prediction of the arrival rate or any relevant success metric based on the route and neighborhood features without running the simulation. One possibility would be the usage of deep learning.

References

1 David Amores, Egemen Tanin, and Maria Vasardani. A proactive route planning approach to navigation errors. International Journal of Geographical Information Science, 35(6):1094–1130, 2021. arXiv:https://doi.org/10.1080/13658816.2020.1820508, doi:10.1080/13658816.2020.1820508.

2 Lorena A. Barba. Terminologies for reproducible research. CoRR, abs/1802.03311, 2018. URL: http://arxiv.org/abs/1802.03311, arXiv:1802.03311.

3 Chris Brunsdon and Alexis Comber. Opening practice: supporting reproducibility and critical spatial data science. Journal of Geographical Systems, 23(4):477–496, Oct 2021. doi:10.1007/s10109-020-00334-2.

4 Paolo Fogliaroni, Dominik Bucher, Nikola Jankovic, and Ioannis Giannopoulos. Intersections of Our World. In Stephan Winter, Amy Griffin, and Monika Sester, editors, 10th International Conference on Geographic Information Science (GIScience 2018), volume 114 of Leibniz International Proceedings in Informatics (LIPIcs), pages 3:1–3:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. URL: http://drops.dagstuhl.de/opus/volltexte/2018/9331, doi:10.4230/LIPIcs.GISCIENCE.2018.3.

5 Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. Gazenav: Gaze-based pedestrian navigation. In Proc of the 17th Int Conf on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '15, page 337–346. ACM, 2015.

6 Klaus Gramann, Paul Hoepner, and Katja Karrer-Gauss. Modified navigation instructions for spatial navigation assistance systems lead to incidental spatial learning. Frontiers in Psychology, 8(FEB), 2017.

7 Peng Han, Jin Wang, Di Yao, Shuo Shang, and Xiangliang Zhang. A graph-based approach for trajectory similarity computation in spatial networks. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery amp; Data Mining, KDD '21, page 556–564, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3447548.3467337.

8 Haosheng Huang, Thomas Mathis, and Robert Weibel. Choose your own route – supporting pedestrian navigation without restricting the user to a predefined route. Cartography and Geographic Information Science, 49(2):95–114, 2022. arXiv:https://doi.org/10.1080/15230406.2021.1983731, doi:10.1080/15230406.2021.1983731.

9 Peter Kedron, Wenwen Li, Stewart Fotheringham, and Michael Goodchild. Reproducibility and replicability: opportunities and challenges for geospatial research. International Journal of Geographical Information Science, 35(3):427–445, 2021. arXiv:https://doi.org/10.1080/13658816.2020.1802032, doi:10.1080/13658816.2020.1802032.

10 Markus Konkol, Christian Kray, and Max Pfeiffer. Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study. International Journal of Geographical Information Science, 33(2):408–429, 2019. arXiv:https://doi.org/10.1080/13658816.2018.1508687, doi:10.1080/13658816.2018.1508687.

11 Saskia Kuliga, Panagiotis Mavros, Martin Brösamle, and Christoph Hölscher. Comparing human wayfinding behavior between a real, existing building, a virtual replica, and two architectural redesigns. In Jur´gis Šk, ilters, Nora S. Newcombe, and David Uttal, editors, Spatial Cognition XII, pages 160–179, Cham, 2020. Springer International Publishing.

12 Carol A. Lawton and Janos Kallai. Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. Sex Roles, 47(9):389–401, Nov 2002. doi:10.1023/A:1021668724970.

13 Bartosz Mazurkiewicz, Markus Kattenbeck, and Ioannis Giannopoulos. Navigating Your Way! Increasing the Freedom of Choice During Wayfinding. In Krzysztof Janowicz and Judith A. Verstegen, editors, 11th International Conference on Geographic Information Science (GIScience 2021) - Part II, volume 208 of Leibniz International Proceedings in Informatics (LIPIcs), pages 9:1–9:16, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. URL: https://drops.dagstuhl.de/opus/volltexte/2021/14768, doi:10.4230/ LIPIcs.GIScience.2021.II.9.

14 Bartosz Mazurkiewicz, Markus Kattenbeck, Peter Kiefer, and Ioannis Giannopoulos. Not Arbitrary, Systematic! Average-Based Route Selection for Navigation Experiments. In Krzysztof Janowicz and Judith A. Verstegen, editors, 11th International Conference on Geographic Information Science (GIScience 2021) - Part I, volume 177 of Leibniz International Proceedings in Informatics (LIPIcs), pages 8:1–8:16, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz Zentrum für Informatik. URL: https://drops.dagstuhl.de/opus/volltexte/2020/13043, doi:10.4230/LIPIcs.GIScience.2021.I.8.

15 Daniel R. Montello and Danqing Xiao. Linguistic and cultural universality of the concept of sense-of-direction. In Max Egenhofer, Nicholas Giudice, Reinhard Moratz, and Michael Worboys, editors, Spatial Information Theory, pages 264–282, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

16 Daniel Nüst, Carlos Granell, Barbara Hofer, Markus Konkol, Frank O. Ostermann, Rusne Sileryte, and Valentina Cerutti. Reproducible research and giscience: an evaluation using agile conference papers. PeerJ, 6:e5072, July 2018. doi:10.7717/peerj.5072.

17 Frank O. Ostermann, Daniel Nüst, Carlos Granell, Barbara Hofer, and Markus Konkol. Reproducible Research and GIScience: An Evaluation Using GIScience Conference Papers. In Krzysztof Janowicz and Judith A. Verstegen, editors, 11th International Conference on Geographic Information Science (GIScience 2021) - Part II, volume 208 of Leibniz International Proceedings in Informatics (LIPIcs), pages 2:1–2:16, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. URL: https://drops.dagstuhl.de/opus/volltexte/2021/14761, doi:10.4230/LIPIcs.GIScience.2021.II.2.

18 Karl Rehrl, Elisabeth Häusler, Sven Leitinger, and Daniel Bell. Pedestrian navigation with augmented reality, voice and digital map: final results from an in situ field study assessing performance and user experience. Journal of Location Based Services, 8(2):75–96, 2014. doi:10.1080/17489725.2014.946975.

19 S. Robinson, M. Jones, P. Eslambolchilar, R. Murray-Smith, and M. Lindborg. “I Did It My Way”: Moving Away from the Tyranny of Turn-by-Turn Pedestrian Navigation. In Proc. of MobileHCI ’10, pages 341–344, 2010.

20 Gian-Luca Savino, Niklas Emanuel, Steven Kowalzik, Felix Kroll, Marvin C. Lange, Matthis Laudan, Rieke Leder, Zhanhua Liang, Dayana Markhabayeva, Martin Schmeißer, Nicolai Schütz, Carolin Stellmacher, Zihe Xu, Kerstin Bub, Thorsten Kluss, Jaime Maldonado, Ernst Kruijff, and Johannes Schöning. Comparing pedestrian navigation methods in virtual reality and real life. In 2019 International Conference on Multimodal Interaction, ICMI ’19, page 16–25, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3340555.3353741.

21 Gian-Luca Savino, Laura Meyer, Eve Emily Sophie Schade, Thora Tenbrink, and Johannes Schöning. Point me in the right direction: Understanding user behaviour with as-the-crow-flies navigation. In 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services, pages 1–11, 2020.

22 J. Thompson, M. Stevenson, and J. S. Wijnands et al. (8). A global analysis of urban design types and road transport injury: an image processing study. The Lancet Planetary Health, 4(1):e32–e42, 2020.

23 Kevin Toohey and Matt Duckham. Trajectory similarity measures. SIGSPATIAL Special, 7(1):43–50, May 2015. doi:10.1145/2782759.2782767.

Pleaese find this paper in the following DOI

Nina Wiedemann, Henry Martin, Esra Suel, Ye Hong & Yanan Xin (2023) Influence of tracking duration on the privacy of individual mobility graphs, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2239190

Please find this paper in the following DOI - Rui Li (2023) Augmented reality landmarks on windshield and their effects on the acquisition of spatial knowledge in autonomous vehicles, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2238661

Please find this paper in the following DOI

Kenzo Milleville, Samuel Van Ackere, Jana Verdoodt, Steven Verstockt, Philippe De Maeyer & Nico Van de Weghe (2023) Exploring the potential of social media to study environmental topics and natural disasters, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2238663

Please find this paper in the following DOI

Irma Kveladze, Marina Georgati, Carsten Kessler & Henning Sten Hansen (2023) Analytics of historical human migration patterns: use cases of Amsterdam and Copenhagen, Journal of Location Based Services, DOI: 10.1080/17489725.2023.2238658

An analysis of potential spaces for implementing geofences in a dynamic bike-sharing system

TIMO HECHEMER[1], JUKKA KRISP[2] AND ANDREAS KELER[3]

[1]Institute of Geography • University of Augsburg • Alter Postweg 118 • 86159 Augsburg

Tel.: +49 151 521 - 50280 • Email: timo.hechemer@student-uni.augsburg.de

[2]Applied Geoinformatics, Institute of Geography • University of Augsburg • Alter Postweg 118 • 86159 Augsburg

Tel.: +49 821 598 - 2756 • Email: jukka.krisp@geo.uni-augsburg.de

[3]Chair of Traffic Engineering and Control • Technical University of Munich • Arcisstraße 21 • 80333 Munich

Tel.: +49 (89) 289 - 22468 • Email: andreas.keler@tum.de

Keywords: Geofence, GIS, Bike-Sharing, Mobility, Digitalisation

Summary: *This paper focuses on the analysis of bike-sharing data in Munich from 2016 - 2019 and discusses its spatial distribution in its central area for further implementation of dockless bikesharing. New developments in transport alternatives can change user behavior and increase the competition between different types of mobility. Since 2017 bike-sharing started to decrease with the launch of electric scooters which don't rely on docks at all. On the one hand it means for Munich's central area that public space will now contain more dangerous obstacles if both - electric scooters and bikes - are highly distributed spatially. On the other hand bike-sharing seems to be in need of an increase in its user flexibility by offering a greater variety of storing the bike which can be achieved by Location Based Services which are already used by electric scooters. Following these thoughts, an implementation of geofences is discussed in order to reduce the spatial distribution over Munich's public space and thus reducing the bikes potential of becoming a dangerous obstacle for other road users.*

Introduction

The bike has always been an important way of getting from one point to another in our modern society and it is hard to imagine a world without it. The reason for traveling doesn't matter – it is used in time of leisure, to get to work or for a short ride to the next shopping district. Thus, the bike is often times used at first choice because it offers independence and can be held at low cost which lie mostly in repairing and changing wheels. As a positive side effect bike users benefit from a lot of healthy aspects – for example getting fresh air and movement that is gentle on the joints which is great for the elderly users. In addition to that, bike users don't rely on parking slots to be able to get anywhere in the city.

So the question arises if existing bike-sharing with docking-stations can still be improved to offer more flexibility. The most direct case would be to not be reliant on a dock at all to store the bike in. This is also the research thematic of this paper which is centered around the spatial independence of bike-sharing with dockless models. Space and time depend on eachother and so moving through space takes time – especially in modern cities that are full of obstacles that can be avoided.

The aim of this paper is to provide useful information about Munich's bike-sharing data that can be used to implement dockless bike-sharing in its central area.

The key questions being answered are...

[1] What is geofencing and how can it be implemented in useful ways?

[2] What does Munich's bike-sharing data tell about its spatial distribution?

[3] How can Munich's central area benefit from a dockless bike-sharing system?

Methods

In this paper the Origin-Destination (OD) data of bike-sharing in Munich from the years 2016 to 2019 is analyzed using ArcGis. The OD contains information about the Starting and Ending position of the ride as well as the rental station. The data that was available through a CSV was imported into ArcGis. The OD will be analyzed using a *Point Clustering Method* that was mentioned earlier and an implementation of geofences for a dockless bike-sharing in Munich's central area will be discussed based on these results.

Results

The results of this Paper contain visualisations of the OD Data by using tools of ArcGis. In Figure 1 Thiessen-Polygons for each Station in Munich's central area were created which have a bar chart for the total number of points of bike-sharing usage in each year (2016 - 2019). It is clearly visible that the need for bike-sharing suffered a great decrease in the year 2017 which only slowly started to recover afterwards. This decrease might have a direct link to the launch of electric scooters in Munich in the year 2017.

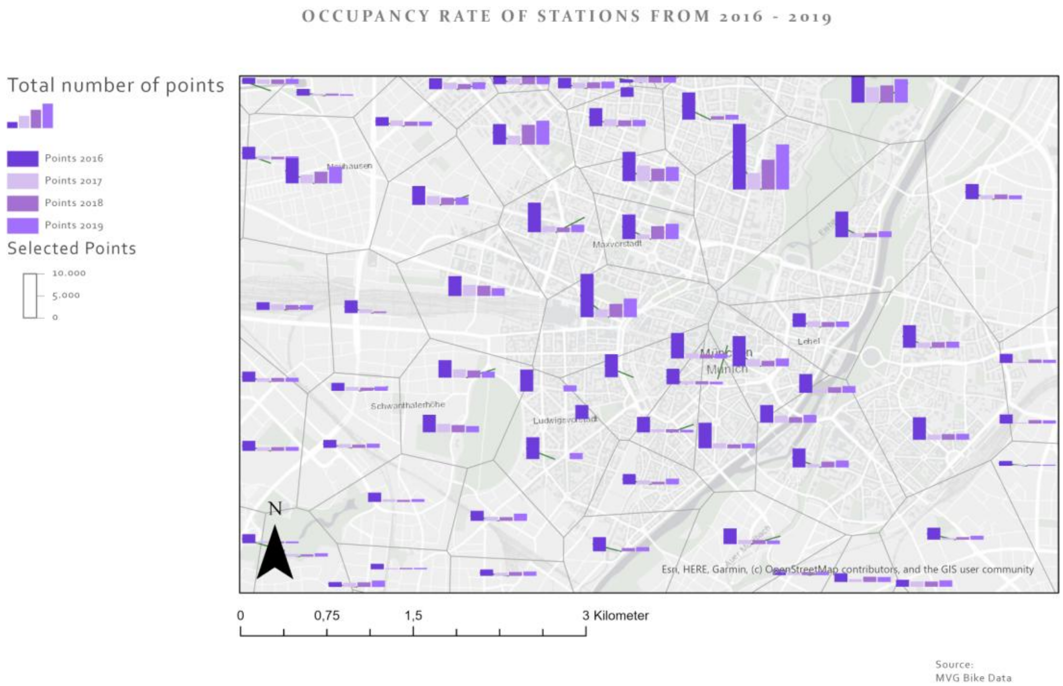


Fig. 1 Occupancy rate of stations between 2016 and 2019

Conclusion

The bike-sharing data of Munich's central area has shown that bike usage is highly distributed over space. This might make it difficult for bike-sharing to reach its full potential of flexibility as stated by Chen et al. (2019: 334). There has been no further development of actual geofences for Munich's central area because it is lacking space for parking bikes in general. Most space is been taken by buildings, streets or public places. The later might seem to offer lots of space but is actually used by people to relax and meet each other which means bikes might become an obstacle in that area. So the increase for bike-sharing that is visible in figure 1 might make geofences through *Location Based Services* interesting for further developement.

References

Chen Z., van Lierop D., Ettema D. (2019): Dockless bike-sharing systems: what are the implications? In: Transportation Reviews 40(3), 333 - 353.

DOI: 10.1080/01441647.2019.1710306

He B., Zhang Y., Chen Y., Gu Z. (2018): A Simple Line Clustering Method for Spatial Analysis with Origin-Destination Data and Its Application to Bike-Sharing Movement Data. In: International Journal of Geo-Information 7 (203), 1 - 16.

Krinkin K., Yudenok K. (2013): Geo-coding in Smart Environment: Integration Principles of Smart-M3 and Geo2Tag. In: Balandin S., Andreev S., Koucheryavi Y. (Eds.): Internet of Things, Smart Spaces, and Next Generation Networking. Berlin, Heidelberg: Springer Verlag, 107 - 116.

Namiot D. (2013): GeoFence Services. In: International Journal of Open Information Technologies 1(9).

Namiot D., Sneps-Sneppe M. (2013): Geofence and Network Proximity. In: Balandin S., Andreev S., Koucheryavi Y. (Eds.): Internet of Things, Smart Spaces, and Next Generation Networking. Berlin, Heidelberg: Springer Verlag, 117 - 127.

Rahate S. W., Shaikh M. Z. (2016): Geo-fencing Infracstructure: Location Based Service. In: International Research Journal of Engeneering and Technology 3(11), 1095 – 1098.

Raviv T., Tzur M., Forma I. A. (2013): Static repositioning in a bike-sharing system: models and solution approaches. In: EURO J Transp Logist 2, 187 - 229.

DOI 10.1007/s13676-012-0017- 6.

Reiss S., Bogenberger K. (2016): Validation of a Relocation Strategy for Munich's Bike Sharing System. In: Transportation Research Procedia 19, 341 – 349.

Sheppard J. K., McGann A., Lanzone M., Swaisgood R. R. (2015): An autonomous GPS geofence alert system to curtail avian fatalities at wind farms. In: Animal Biotelemetry 3(43).

An Emerging Conceptual Model for Curating Engaging Leisure Walking Recommendations

James Williams[1], James Pinchin [1], Adrian Hazzard [2], Gary Priestnall [3], Stefano Cavazzi [4] and Andrea Ballatore [5]

[1] Nottingham Geospatial Institute • University of Nottingham • Nottingham • United Kingdom

E-Mail: James.Williams@Nottingham.ac.uk, James.Pinchin@Nottingham.ac.uk

[2] Mixed Reality Lab • University of Nottingham • Nottingham • United Kingdom

E-Mail: Adrian.Hazzard@Nottingham.ac.uk

[3] School of Geography • University of Nottingham • Nottingham • United Kingdom

E-Mail: Gary.Priestnall@Nottingham.ac.uk

[4] Ordnance Survey

E-Mail: Stefano.Cavazzi@os.uk

[5] Department of Digital Humanities • King's College London • London • United Kingdom

E-Mail: Andrea.Ballatore@kcl.ac.uk

Keywords: Leisure Walking, Location Based Services, Route Recommendation, Mobile Geospatial Computing

Summary: *Providing routes to leisure walkers requires alternative recommendation scenarios to those used in tourism routing systems. In this paper, we present an emerging conceptual model of three scenarios for curating leisure walking route recommendations. Our recommendation scenarios consider the highest ranked similar walks, routes for new application users, and a progressively changing route recommendation scenario. Conceptual models for these scenarios are presented and the challenges in completing this research are considered. Feedback received on these early conceptual models will be used to further design a recommendation framework for curating engaging leisure walking experiences.*

Introduction

Leisure walking is an outdoor activity that can be undertaken for the purposes of getting outside, wellbeing, and exploring new places (Williams et al., 2021). Providing rich, contextual, and interesting walks for individuals is therefore a unique challenge that has been considered in previous literature. Watts & Bauer (2022) investigate the design and implementation of peaceful walks using a rating prediction tool that considers noise and natural features. Quercia et al. (2014) reports on walking route recommendations in the city, providing participants with routing algorithms that attempt to provide short, beautiful, quiet, and happy walks from crowdsourced perception data. Providing personalised recommendations at scale presents a problem in curating new routes, especially when considering research beyond that of moving between the most popular points of interest (POIs) (e.g., Gavalas et al., 2017), or outside of urban areas. We use the term curated to refer to the selection and organisation of POIs to create route experiences which meet the needs of leisure walkers.

The conceptual model presented in this paper is an emerging set of recommendation scenarios that will be further developed in the rest of our research project. The remainder of this paper presents the current recommendation scenarios, the challenges in relation to providing recommendations, and the expected outcomes and future work for the research.

Recommendation Scenarios

Based on the identified gap in applications and literature, we designed three initial conceptual models for a leisure walking route recommendation system. The system proposed is designed to take a hybrid approach to providing recommendations, making use of both content-based and collaborative filtering methods (Aggarwal, 2016). The emerging system design was proposed in a way that multiple scenarios could make use of the same implicit and explicit input interaction (Ballatore & Bertolotto, 2015) to personalise application content. Figure 1 presents a high-level design diagram of this proposed model.

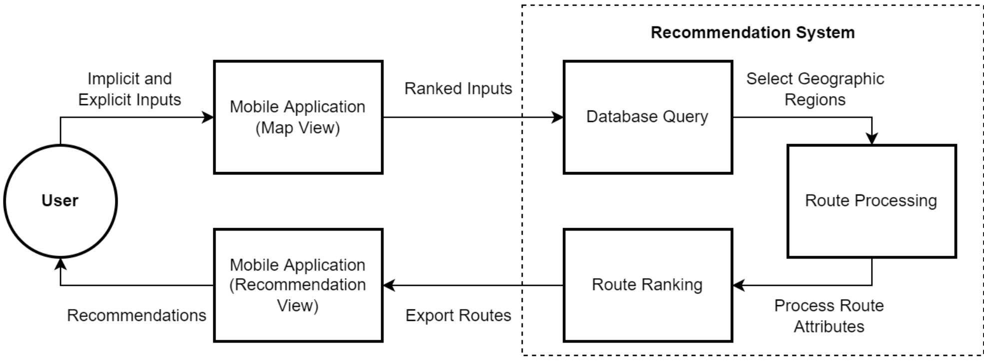


Fig. 1: A high-level overview of the conceptual model for the proposed recommendation system.

We propose that each recommendation scenario can also apply to two types of routing, defined as:

- **Dynamically Generated Routes:** Custom routes that are generated using existing attributes from datasets such as POIs or mobility traces, routing is then performed between each instance.
- **Pre-defined Routes:** Routes that have been curated or uploaded by users, this may be user generated content making use of ambient or volunteered geographic information.

Top Ranked Similar Routes

The first proposed scenario introduces the top-ranked similar walks to the application user. This proposed scenario uses ranked inputs from the user and stores these rankings in the system, when a new request is received the database is queried for similar route attributes (e.g., POIs, features). We propose that these routes are then ranked by the most popular in the selected geographic area and presented to the user.

Routes for New Users

The second scenario considers the display of routes to new application users, who may not have large amounts of information already in the system (e.g., the cold start problem). This scenario ranks known context (e.g., geolocation) and known responses to onboarding questions (e.g., where do you like walking?). The system will then query based upon this data and generate a distinct set of routes.

Progressive Route Strategies

The final proposed scenario presents a progressively changing selection of routes, enabling user selected strategy requirements to be supported through route recommendations. For example, a user may look to increase route complexity over a specified time period, meaning a plan is stored within the application and used to process and then present these recommendations.

Research Challenges

Some challenges exist in providing contextual and interesting route recommendations for leisure walkers, including:

Lack of Data. A lack of public data that exists in regards to natural or more subjective places along a walk make it difficult to curate leisure walking recommendations. Tourism-based research can use Foursquare POIs (e.g., Yang et al., 2015) to apply attributes to routes, however, this approach is not possible in more rural areas due to an absence of identified physical locations.

Subjective Data. Data relating to leisure walking shares similar problems to that of other types of user generated content. It can be considered that user data required to apply context to routes needs to be captured through implicit or explicit interactions, requiring considerations from designers as to the privacy, reliability, and scalability of this data.

Expected Outcomes and Future Work

The emerging conceptual models presented are expected to be used within a framework for the curation of leisure walking route recommendations. We hope to understand how users engage with leisure walking routes through conducting a user study. With plans to investigate the use of platial information, a type of information relating to place as opposed to spatial representations (Westerholt et al., 2018). Our aim of this will be to investigate how platial information can be identified from meaningful interactions and used in the curation of new routes.

Acknowledgements

James Williams is supported by the Horizon Centre for Doctoral Training at the University of Nottingham (UKRI Grant No. EP/S023305/1) and by the external partner Ordnance Survey.

References

Aggarwal, C.C. 2016: Ensemble-Based and Hybrid Recommender Systems. In: Aggarwal, C.C. (Eds.): Recommender Systems. – Springer, pp. 199-224. doi:10.1007/978-3-319-29659-3_6

Ballatore, A. & Bertolotto, M. 2015: Personalizing Maps. – Communications of the ACM 58: 68-74. doi:10.1145/2756546

Quercia, D., Schifanella, R. & Aiello, L.M., 2014: The Shortest Path to Happiness: Recommending Beautiful, Quiet, and Happy Routes in the City – Proceedings of the 25th ACM conference on Hypertext and social media: 116-125. doi:10.1145/2631775.2631799

Gavalas, D., Kasapakis, V., Konstantopoulos, C., Pantziou, G. & Vathis, N., 2017: Scenic route planning for tourists. – Personal and Ubiquitous Computing 21: 137-155. doi:10.1007/s00779-016-0971-3

Watts, G. & Bauer, J., 2022: Tranquillity trails – design, implementation and benefits for healthy leisure. – World Leisure Journal 64: 156-165. doi:10.1080/16078055.2021.1927165

Westerholt, R., Mocnik, F.-B. & Zipf, A. 2018: Introduction to the PLATIAL 18 Workshop on Platial Analysis. – First Workshop on Platial Analysis (PLATIAL'18): 1-5. doi:10.5281/zenodo.1475267

Williams, J., Pinchin, J., Hazzard, A. & Priestnall, G., 2021: Survey of Leisure Walking Behaviours and Activity Tracking Use: Emerging Themes and Design Considerations. In: Basiri, A., Gartner, G. & Huang, H. – Proceedings of the 16th International Conference on Location Based Services (LBS 2021): 136-141. doi:10.34726/1760

Yang, D., Zhang, D., Chen, L. & Qu, B., 2015: NationTelescope: Monitoring and visualizing large-scale collective behavior in LBSNs. – Journal of Network and Computer Applications 55: 170-180. doi:10.1016/j.jnca.2015.05.010

Implementation of an Indoor Positioning System (IPS) on a university campus

JOSHUA PORZLER, RAINER SCHÄFFNER AND JAN WILKENING
University of Applied Sciences Würzburg-Schweinfurt • Röntgenring 8 • 97072 Würzburg, Germany
Tel.: +49 (0)93 3511-8229 • E-Mail: jan.wilkening@fhws.de

Introduction

Getting lost in a building is a common experience, especially for first-time visitors. This also applies to university campuses. While finding the current position *outdoors* with smartphone and GNSS sensors is straightforward, *indoor* positioning is a bigger challenge (Kunhoth et al., 2020, Koyuncu & Yang, 2010, Farid et al., 2013). Since most routing tasks start at the current position and most targets on a campus are indoors (lecture rooms, offices, toilets or refectories), we were interested in investigating the possibilities to implement an Indoor Positioning System (IPS) at our university campus. The goal was to show the "blue dot" with the current indoor position on mobile devices as accurately as possible.

Evaluating indoor positioning technologies

In this study, we first examined different technologies for the implementation of Indoor Positioning Systems (IPS). These technologies included ultra-wide band (UWB), radio-frequency identification (RFID), wireless fidelity (WiFi), ultrasound, and Bluetooth Low Energy (BLE).

For our project, we decided in favor of BLE, because it offered the best ratio between accuracy and costs. A similar setup was implemented and tested in other studies (e.g., Satan, 2018, Lee et al., 2018) As a software component, we chose ArcGIS Indoors, since the university's Campus Information System had already been built based on ArcGIS Online and ArcGIS Enterprise, and we could use ArcGIS Indoors without any additional cost.

Implementation of an IPS

In a first step, Bluetooth Low Energy beacons and the ArcGIS Indoors technologies were tested in a residential building as a prototype, since on-campus experiments could not be conducted due to the ongoing Covid-19 pandemic.

The next step consisted in the planning of the beacon distribution in Desktop GIS (ArcGIS Pro, see Figure 1). After completion of the planning stage, the ArcGIS Indoors model was prepared. We then placed 85 beacons in the buildings, mostly in corridors and stairways, and thus covered 4000 m² in different parts of the campus. When placement and labeling were completed, we recorded beacon signals with a third-party app (indoo.rs) using SLAM (Simultaneous Localization and Mapping) technology.

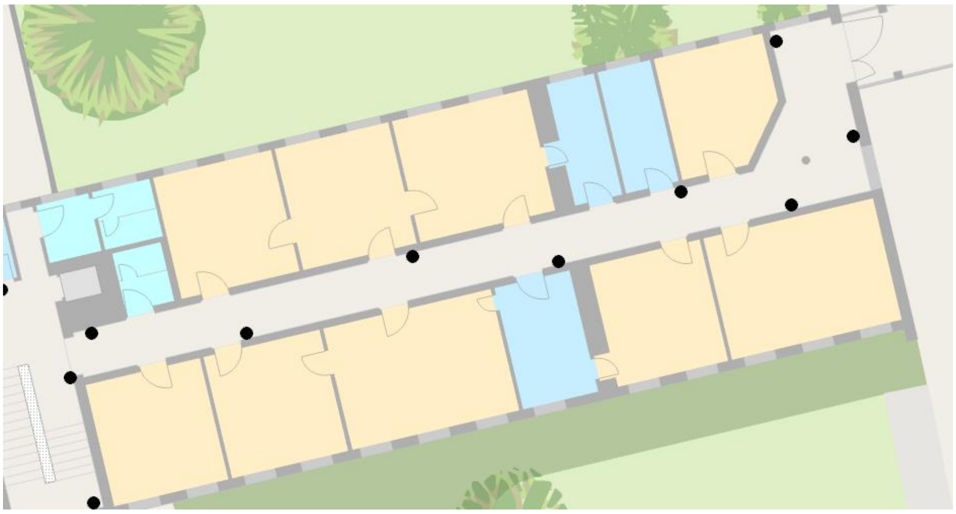


Fig. 1: Beacon placement in ArcGIS Pro

In initial on-campus experiments, we achieved a positional accuracy between one and six meters. After consolidation, change of the beacons' position and power strength, we could improve the accuracy to values between one and three meters (see Figure 2).

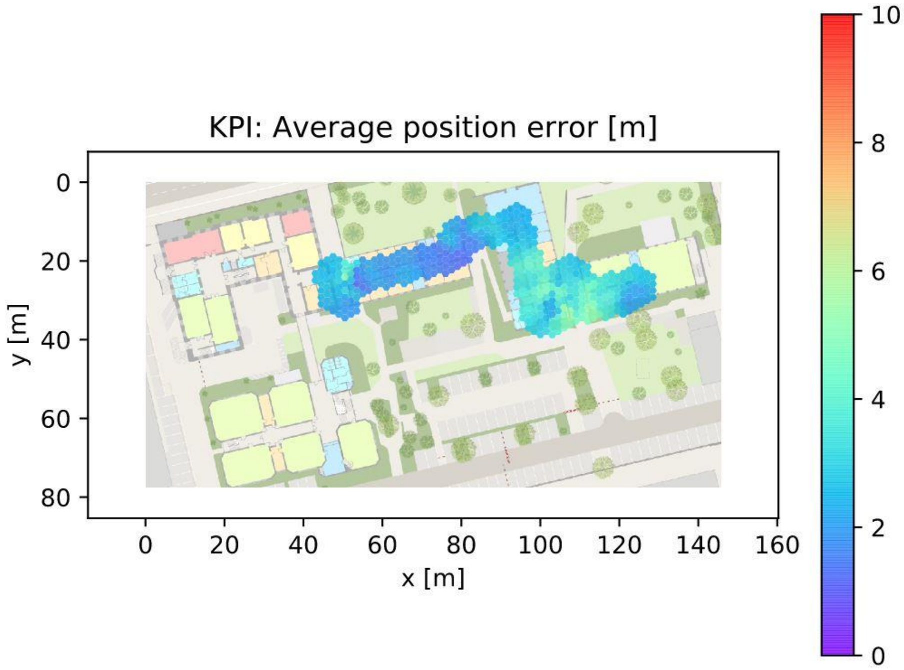


Fig. 2: Average positioning errors in one building

Due to the different structure of the five buildings on the campus, some conclusions for the improvement of the beacon placements were drawn. For instance, we achieved lower levels of accuracy in lower levels of the buildings, due to thicker walls, low ceilings and many curves and edges. On these levels, we had to place more beacons at a lower height, because obstacles tended to be close to the ceiling. In long aisles with more beacons in other areas, accuracy was higher, due to higher ceilings and a fewer amount of obstacles for signals. In general, long aisles inside a building required a larger number of beacons per m² than quadratic lecture rooms.

In some cases, it also made sense to reduce the signal power, because with a higher power, the "blue dot" with the current position on the map (see Figure 3) sometimes "jumps around".

Fig. 3: „Blue dot" on a mobile device after successful IPS implementation

After achieving satisfying levels of accuracy in all public parts of the buildings where the beacons had been placed, user experiments were conducted with different Android and iOS devices. For this part of the experiment, 30 measurement points were marked on the ground, and participants recorded the "blue dot". Based on these recordings, we could analyze the deviations between the blue dot and the measurement points more precisely and further calibrate the beacon distribution.

Conclusion and future work

We successfully implemented an IPS at our university campus, and the results can be displayed with mobile devices. We achieved positioning accuracies of one to three meters inside buildings, by changing the initial configuration (such as signal power and the density) and measuring accuracy results.

Future work should investigate the factors that influence the positional accuracy in buildings. Furthermore, IPS can be expanded to the entire campus and should be enhanced with a routing and tracking function. In the meantime, new products (such as ArcGIS IPS) have arrived on the market, which further facilitate combining indoor positioning systems with GIS technology. These products also allow users with less GIS knowledge to implement IPS in bigger buildings more efficiently.

The topic of indoor positioning has an enormous potential for GIScience, since indoor positioning is an important issue for many types of facilities, such as airports, hospitals, railway stations, large industry complexes, museums or libraries.

References

FARID, Z., NORDIN, R. & ISMAIL, M., 2013: Recent advances in wireless indoor localization techniques and system. Journal of Computer Networks and Communication 2013:12.

KOYUNCU, H. & YANG, S.H., 2010: A survey of indoor positioning and object locating systems. International Journal of Computer Science and Network Security 10 (5):121–128.

KUNHOTH, J., KARKAR, A., AL-MAADEED, S. & AL-ALI, A., 2020: Indoor positioning and wayfinding systems: A survey. Human-centric Computing and Information Sciences 2020:10-18.

LEE, K., NAM, Y. & MIN, S.D., 2018: An indoor localization solution using bluetooth RSSI and multiple sensors on a smartphone. Multimedia Tools and Applications 77(10):12635–12654.

SATAN, A, 2018: Bluetooth-based indoor navigation mobile system. In: 2018 19th international carpathian control conference (ICCC): 332–337.

Understanding the Uncertainty in Measuring Close Contacts Using Mobile Phone Location Data

SONG GAO

Department of Geography • University of Wisconsin-Madison• 550 N Park St • WI 53706 USA

E-Mail: song.gao@wisc.edu

Keywords: Mobile Device Data, COVID-19, GIScience, LBS

Summary: *During the COVID-19 pandemic, great efforts have been made on measuring human mobility patterns and interpersonal close contacts using mobile phone location big data. However, few studies analyzed the uncertainty in those measures. This research aims to address two main uncertainty issues in measuring close contacts using mobile device data: noise and choice of report location. Using large-scale mobile device panel data at the state level in the US, our experiment results show that the noise (false positives) detected in close contacts generate larger errors than the choice of report location (home vs. event location). In addition, the population density (rather than total population) is positively correlated with the close contact measurement errors.*

Introduction

Understanding individuals' mobility patterns and interpersonal close contacts is critical in modeling the spatial spread of infectious epidemics. During the COVID-19 pandemic, great efforts have been made on measuring human mobility patterns and tracing close contacts using mobile phone location big data and understanding their associations with COVID-19 infection rates (Chang et al. 2021; Hou et al. 2022; Crawford et al. 2022). For example, researchers found a positive association between human mobility and COVID-19 infections at the county and the state geographic scales in the US (Gao et al. 2020; Xiong et al. 2020). It is worth noting that reduced mobility doesn't necessarily ensure that people follow the social (physical) distancing guideline "staying at least 6 feet (2 meters) from other people". Due to the mobile phone GPS location horizontal error and uncertainty, such physical distancing patterns cannot be identified from the use of aggregated mobility data. Individual-level interpersonal close-contacts (spatiotemporal co-location patterns) can be estimated using mobile device data and have shown better association with the infection rates than the mobility metrics (Crawford et al. 2022).

However, there are multiple sources of uncertainty in measuring interpersonal close contacts using mobile device location data. (1) the positioning accuracy of mobile phone devices (Zandbergen 2009); (2) the noise (false positives) in the measurement of close contacts proxy (i.e., spatiotemporal co-location events); and (3) the choice of report location (residence vs. events). The positioning accuracy is various given different mobile device types and with different environmental contexts, which is hard to measure. Therefore, this research mainly focuses on the assessment of other two uncertainty sources: noise and choice of report location.

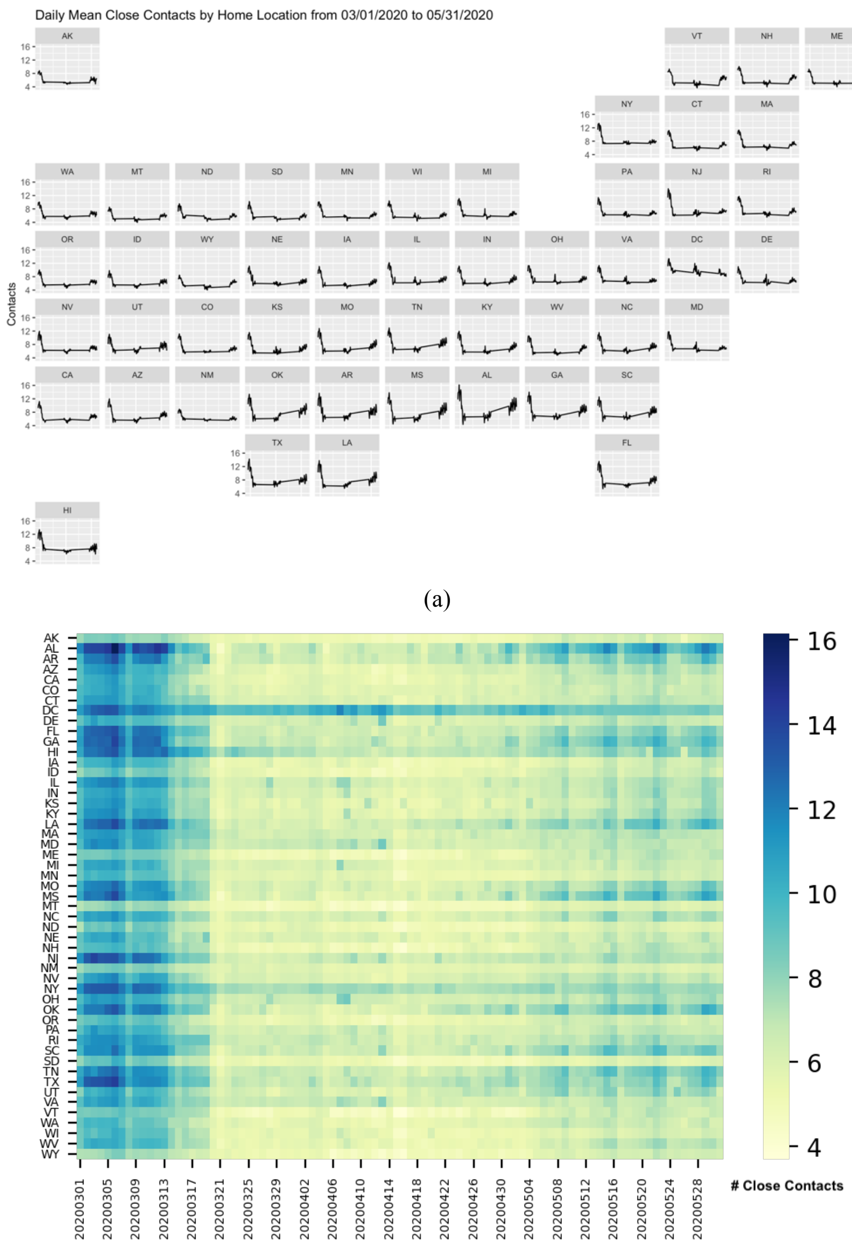
Methods

Close Contacts Measurements

We employ a large-scale anonymized mobile phone device location panel data from UberMedia that covers 70% of the US population. The spatiotemporal co-location events (close contacts proxy) of each device is measured by calculating how many devices come within 5 meters of the target device within 5 minutes for a given time period (daily updates) based on their GPS locations. The first type of uncertainty comes from the spatial context; many detected "close contacts" are along major roadways and may be in separate vehicles, which are false positives from the epidemic modeling perspective and should be filtered as noise. The second type of uncertainty comes from the choice of report location (home vs. event location) when aggregating individual device data into a geographic unit (e.g., census block, tract, zip code, county, and state). The close contacts pattern of a region could be different when changing the report location of a device from its common evening location ("*home*") from 8pm to 8am on weekdays to the *event location* where the spatiotemporal co-location events are detected and may occur in non-residence locations such as in work and social places.

Evaluation Metrics

In order to assess the uncertainty impacts of two sources on detected close contact patterns, we choose two commonly used evaluation metrics: root-mean-square-error (RMSE) and mean-absolute-error (MAE) to compare different processing methods on the state-level daily close contact measurements using the mobile device panel data over the study period.



(a)

(b)

Fig. 1: The temporal patterns of daily mean close contacts (filter noise) (a) time series panel by each state in the US; and (b) its corresponding heatmap.

Results

We first compute the US state-level daily mean close contacts of each device by home location and filter noise by removing all the co-location data points with moving speed of 25 km/hr as the baseline. Fig.1 shows the temporal patterns and heatmap of daily mean close contacts by each state in the US from March 1st to May 30, 2020. We can see that the overall downward close contact trend during the state lockdown period from March 19th (when California was the first state issued the stay-at-home order) to the beginning of May, then the upward trend bounced back in the Memorial Day week (late May) when more people went out. However, for Washington D.C., it had persistent high close contacts across the study period, which might link to the gathering events in the presidential election year.

Then, we compare the differences on the close contacts measurements from the above-mentioned uncertainty perspectives including four types: (a) home location filtered noise; (b) home location with noise; (c) event location filtered noise; and (d) event location with noise. As shown in Fig.2, their temporal patterns are quite different although the general temporal trend remains during the statewide lockdown period. The median of the RMSE of state-level daily mean close contacts with and without filtering "noise" (false positives) is about 15 across all US states, which is larger than the choice of event location (median RMSE: 1.36). The similar results remain when using the MAE metric, as shown in Fig. 3.

Our experiment results show that the "noise" (false positives) detected in close contacts generates larger errors than the choice of report location (home vs. event location). In addition, the population density (rather than total population) is positively correlated with the close contact measurement errors (false positives), with Pearson correlation coefficients of 0.46 (p-value < 0.05) for RMSE and 0.47 (p-value < 0.05) for MAE, which provide critical information in understanding the virus spread and guiding social distancing policy in different regions using mobile device location big data.

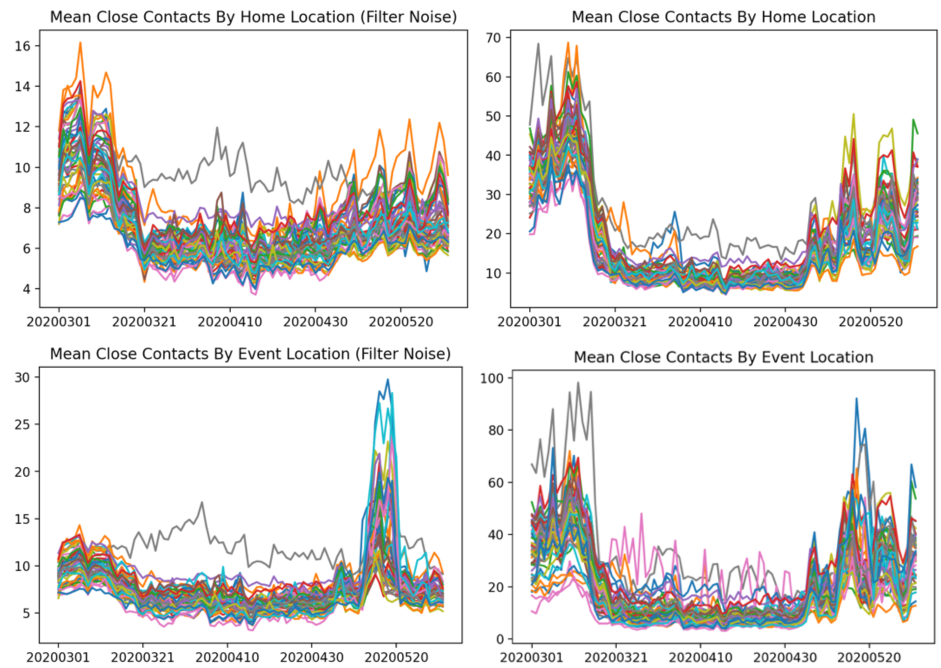


Fig. 2: The state-level temporal patterns of mean close contacts (y-axis) using different approaches.

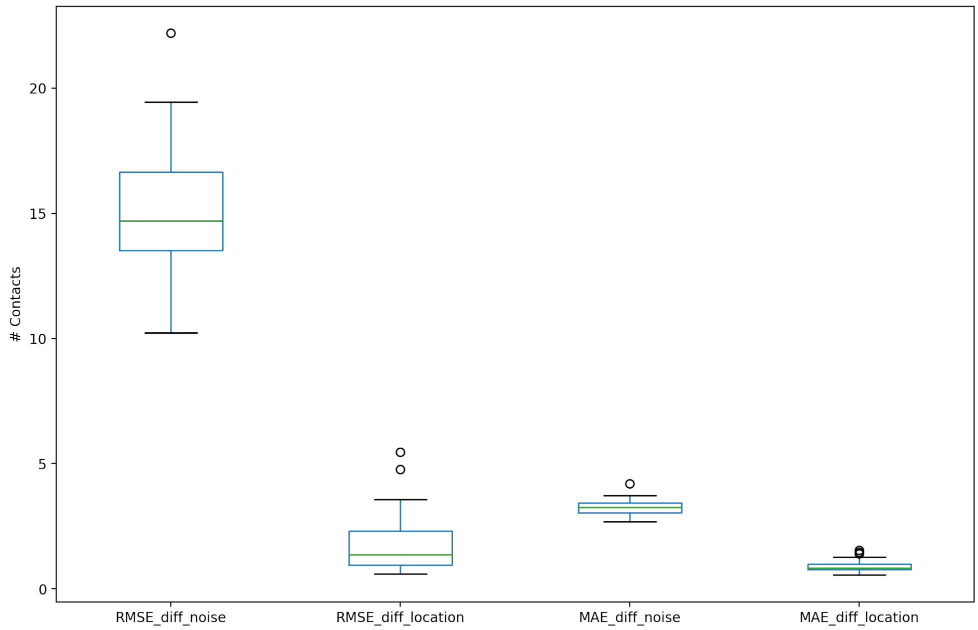


Fig. 3: The RMSE and MAE results for comparing different close contacts measurement methods.

Acknowledgement
The author would like to thank the UberMedia Inc. for providing the anonymous mobile device data and acknowledge the funding support provided by the U.S. National Science Foundation (Award No. BCS-2027375). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Support for this research and travel is provided by the University of Wisconsin - Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.

References
- Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589(7840), 82-87.
- Crawford, F. W., Jones, S. A., Cartter, M., Dean, S. G., Warren, J. L., Li, Z. R., ... & Morozova, O. (2022). Impact of close interpersonal contact on COVID-19 incidence: Evidence from 1 year of mobile device data. *Science Advances*, 8(1), eabi5499.
- Gao, S., Rao, J., Kang, Y., Liang, Y., Kruse, J., Dopfer, D., ... & Patz, J. A. (2020). Association of mobile phone location data indications of travel and stay-at-home mandates with COVID-19 infection rates in the US. *JAMA Network Open*, 3(9), e2020485-e2020485.
- Hou, X., Gao, S., Li, Q., Kang, Y., Chen, N., Chen, K., ... & Patz, J. A. (2021). Intracounty modeling of COVID-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proceedings of the National Academy of Sciences*, 118(24).
- Xiong, C., Hu, S., Yang, M., Luo, W., & Zhang, L. (2020). Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. *Proceedings of the National Academy of Sciences*, 117(44), 27087-27089.
- Zandbergen, P. A. (2009). Accuracy of iPhone locations: A comparison of assisted GPS, WiFi and cellular positioning. *Transactions in GIS*, 13, 5-25.

Attractivity context graph for exploring the travel activity of Flickr users

MATAN MOR AND SAGI DALYOT

Mapping and Geoinformation Engineering • The Technion • Haifa 3200003 • E-Mail: matam.mor@campus.technion.ac.il

Geotagged photos are uploaded by users to social media photo-sharing online websites, which become very popular and commonly used by travellers to share their experiences. Interpreting these user-generated 'digital footprints' can be used to reconstruct travel trajectories of users to explore their travel activity and knowledge on the urban environment. In this work, Flickr geotagged crowdsource photo database is showcased to differentiate between the travel activity of two user-groups: tourists and locals. We develop an activity context graph based on two attractivity matrices: popular frequently visited places and locations, and popular connectivity between them. By analysing the geotagged photographs of all Flickr users visiting an urban destination, the graph's nodes and edges resemble the overall travel activity patterns of both user groups.

Classifying between the two user groups is non-linear, meaning that class boundaries cannot be well approximated using pre-defined parameters, such as visit duration. Accordingly, we develop a supervised Random Forest machine-learning classification model to differentiate between tourists and locals. The model is based on an array of unique features, which are calculated based on the users' a) social media footprint (e.g., total number of photos the user uploaded to Flickr), b) travel statistics (e.g., the accumulated distance covered by the user according to the sequence of all photos taken in the area), and, c) travel behaviour (city centrality indices that reflect the areas traversed by the user). After classification, for each user-group we construct the respective attractivity graph to analyse the unique travel patterns and behaviours.

Using the Random Forest model, the classification produces very accurate results of F_1-score above 90. The model shows that unique features, such as returning visits, which are not used in tourism research, are important to differentiate between the two groups. Based on the user-group differentiation, we construct the attractivity graph for New-York City, depicted in Fig 1. We show that tourists and locals travel to similar popular destinations. However, locals travel also to more dispersed places, while traversing through more peripheral zones in New-York City. Moreover, tourists prefer to travel over short distances to maximize their travel experience, while locals allow themselves to travel over longer distances. The results show that travel information can be explored for measuring and analysing travel activity of different groups from user-generated content, aimed at enriching existing knowledge related to travel analysis and management in urban areas.

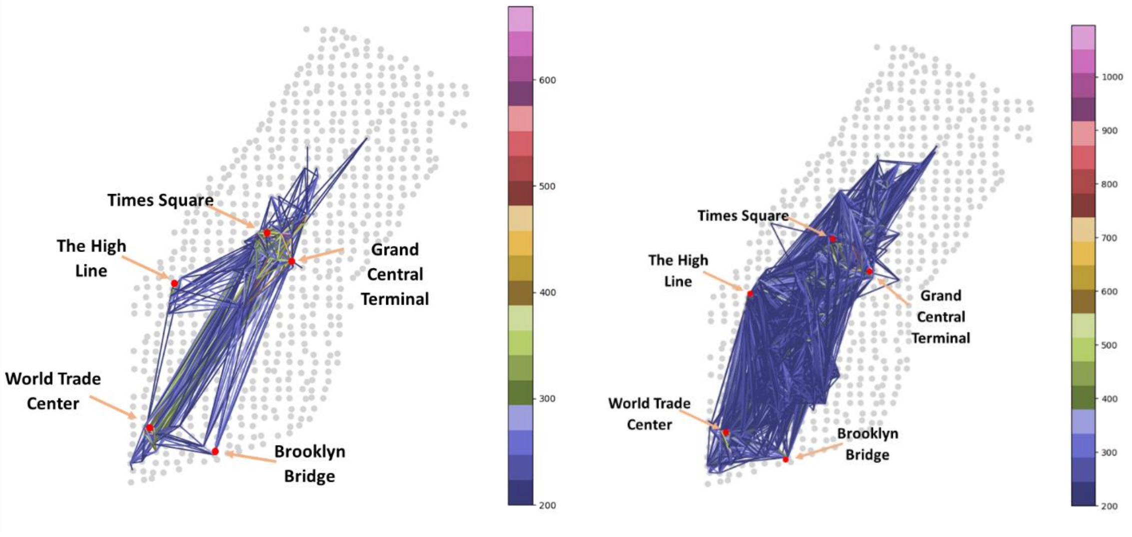


Fig 1. Attractivity graph for New-York City: user visitation connectivity (edges) above the value of 200 between nodes (popular visited places and locations): tourists (left), and locals (right).

Integrating real-time data into a Campus Information System with senseBox and GeoEventServer

JAN ERHARDT, RAINER SCHÄFFNER AND JAN WILKENING

University of Applied Sciences Würzburg-Schweinfurt • Röntgenring 8 • 97072 Würzburg, Germany

Tel.: +49 (0)93 3511-8229 • E-Mail: jan.wilkening@fhws.de

Introduction

GIS-based Campus Information Systems (CIS) can progressively be enhanced with different data and functionalities (Bansal, 2014, Mittlböck et al., 2017, Wilkening et al., 2018, 2019). At the time of writing, the CIS at our university consists of a 2D map and a 3D scene for visualization purposes, a routing network, and a database with information on buildings, lecture halls, offices and other facilities that can be queried. The CIS is under constant development: New features are being added in order to offer students and visitors a broad variety of information.

In this paper, we focus on the integration of real-time environmental data into the CIS. These data include temperature, air pressure, humidity, and particulate matter and are acquired by a sensor unit called "senseBox" (Bröring et al., 2011, see Figure 1).

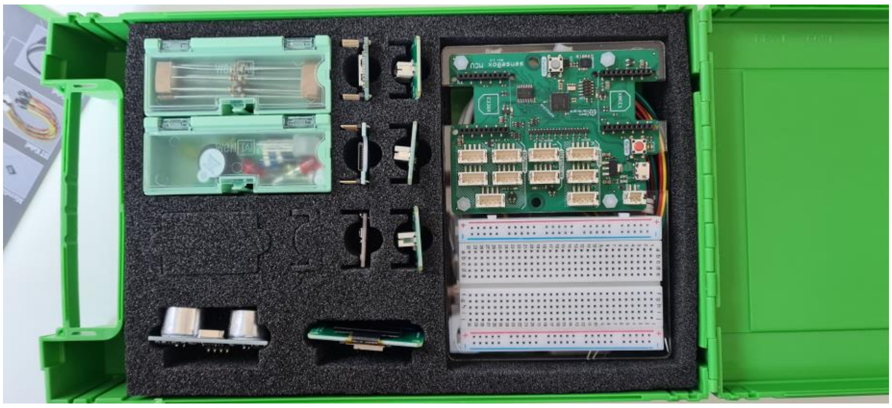


Fig. 1: senseBox unit with interfaces for different sensors

Integrating a stationary senseBox in the university network

In order to connect the senseBox to the internet, one has to place a WiFi Bee, enter WiFi credentials in the C++ code and use Blockly or Arduino IDE to transfer the code to the senseBox. Afterwards, the senseBox can be programmed to send its measurements to the website opensensemap.org.

While integrating the senseBox into a private WLAN in the manner described above was straightforward, the integration of the senseBox to the university's digital infrastructure was a bigger challenge, due to higher security requirements. The WiFi bee thus could not be easily integrated into the university network by simply adding WiFi credentials. The solution consisted of using an upstream router and a mobile hotspot.

After the senseBox was connected to the internet, we used ArcGIS GeoEvent Manager to query the openSenseMap API and process the data. GeoEvent Manager ist a real-time extension to ArcGIS Enterprise, which facilitates Server GIS on-premises with specific web services that cannot be hosted on ArcGIS Online. The data from the senseBox was collected with the input connector "Poll an External Website for JSON" and then forwarded to two output connectors: The first connector updates an existing point feature in a hosted Feature Service to recorded the latest measurements, which are displayed by a pop-up in the campus information system (see Figure 2). The second output connector adds measurements as new features to a table, which is serving as a data archive.

Fig. 2: Position of the senseBox in the CIS and pop-up with attribute values

The next step consisted of preparing a web app (ArcGIS Dashboard) for analysis and visualization purposes. This web app queries the table and visualizes measurements in real-time. The contents of the dashboard include charts with the temporal development of the measured data and a small web map with the position of the senseBox (see Figure 3). The measurements are held up for 24 hours and can be downloaded from the archive table on ArcGIS Online.

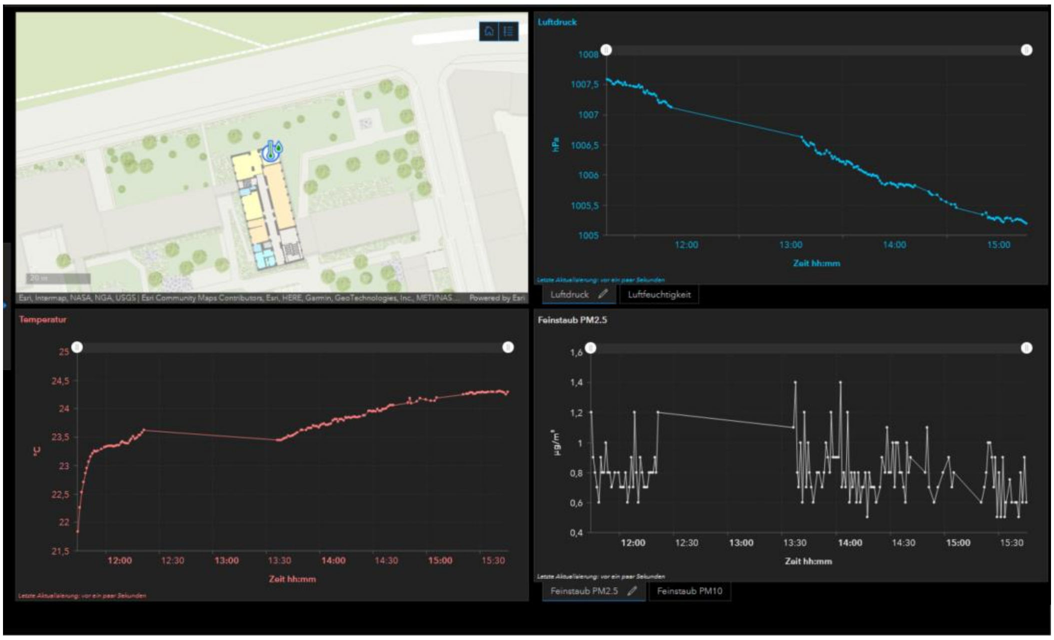


Fig. 3: Interactive dashboard with web map and line charts for attribute values

Mobile senseBox

We also tested the setup of a second (mobile) senseBox, which is equipped with a GNSS sensor. As a mobile unit, it links the captured environmental data to geographic coordinates and displays the live location and the attribute values on a Web Map. While this senseBox is not part of the campus information system yet, it can be used as a tool for easy mobile data capturing.

The mobile senseBox is based on a Stream Service configured in ArcGIS GeoEvent Manager and stored on ArcGIS Server. A Stream Service emphasizes low latency, real-time data dissemination for client-server data flows[1]. The Stream Service, which measures the coordinates and the attributes of the mobile sensebox, was then saved as a Feature Layer on ArcGIS Online to make it accessible for several other apps. The recorded data is stored in a table on ArcGIS Online, similar to the data of the stationary senseBox.

For the mobile senseBox, we also configured a dashboard with the position and charts for the measurements, which is optimized for handheld devices like smartphones and tablets.

Summary and outlook

The project described above shows that it is possible to integrate senseBoxes into Campus Information Systems (CIS) within the scope of a bachelor's thesis. With the means of GeoEvent Server as a software component and its input and output connector capabilities, we also created several Feature Services. These web services provide the basis for different web apps, such as dashboards.

Due to the few amount of tutorials, the integration process consisted of many trial-and-error iterations. It was necessary to simplify and combine input and output processes to reduce data volume, reduce data traffic and achieve a stable configuration.

Currently, other sensors are evaluated that can be integrated en route to a "Smart Campus" at our university. The integration of a senseBox into a CIS is an important step to help stakeholders to make smarter decisions, save energy and optimize resources, based on actual data values that are displayed in a web app.

References

BANSAL, V. K., 2014: Use of Geographic Information Systems in Spatial Planning: A case study of an institute campus. Journal of Computing in Civil Engineering, 28(4), 05014002.

BRÖRING, A., REMKE, A, & LASNIA, D., 2011: SenseBox – A generic sensor platform for the web of things. In: International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services : 186-196. Springer, Berlin, Heidelberg.

MITTLBÖCK, M. KNOTH, L. & VOCKNER, B., 2017: Universitäre Campus Maps – Beispiele aus Österreich und Nordamerika: Status quo & quo vadis? AGIT – Journal für Angewandte Geoinformatik (3-2017), VDE Verlag, Herbert Wichmann Verlag, Berlin: 374-382.

WILKENING, J., KAPAJ, A. & CRON, J., 2019: Creating a 3D Campus Routing Information System with ArcGIS Indoors. Dreiländertagung der DGPF, der OVG und der SGPF in Vienna, Austria –Publikationen der DGPF, Band 28, 2019.

WILKENING, J., SCHÄFFNER, R. & STAUB, T., 2018: Interaktiver 3D-Routenplaner für den Campus Röntgenring in Würzburg. AGIT – Journal für Angewandte Geoinformatik 4-2018:. 35-40.

[1] https://enterprise.arcgis.com/en/server/latest/publish-services/windows/stream-services.htm

Reproducible methods for spatio-temporal accessibility and mobility studies

HENRIKKI TENKANEN [1] AND CHRISTOPH FINK[2]

[1] Department of Built Environment • Aalto University • Otakaari 4 • 02150 Espoo, Finland

[2] Digital Geography Lab • University of Helsinki • Gustaf Hällströminkatu 2 • 00560 Helsinki, Finland

• E-Mail: henrikki.tenkanen@aalto.fi

Keywords: OpenStreetMap, data extraction, accessibility analysis, reproducible methods

Introduction

Availability of open data and powerful open source libraries have significantly altered the way contemporary geospatial science is conducted. For instance, OpenStreetMap (OSM) has become a central data source for geographical analysis, used extensively both in academic research and business. As a free, crowdsourced digital map and database of the world (Haklay and Weber, 2008), it provides crucial geographical data that are commonly used for providing useful services for people, such as navigation, and analyzing various urban phenomena. In addition, datasets such as General Transit Feed Specification (GTFS) provides information about public transportation systems and how people can reach places (Mahajan et al., 2021). GTFS data is available from thousands of cities across the world, allowing the researchers to study accessibility and mobility related questions in urban areas.

In this talk, we introduce a couple of new open source Python libraries called pyrosm and r5py that make it possible to conduct large-scale geospatial analyses in a reproducible manner. Pyrosm (Tenkanen, 2021) is an efficient data extraction tool that enables extracting various useful datasets for accessibility and mobility studies from OpenStreetMap (buildings, roads, services and other points of interest, landuse, etc.). Pyrosm converts OSM data into GeoDataFrame which is the basic data structure of geopandas. Geopandas is the core library for modern geospatial analysis in Python, having a broad ecosystem of spatial analysis libraries built on top of it. Pyrosm has been optimized for performance allowing e.g. large-scale spatial network analysis based OSM data extracts covering city regions or even countries (Figure 1). The library also supports extracting historical data from OSM (OSH.PBF format), which enables conducting spatio-temporal analysis, such as detect changes in urban areas. The historical data extraction functionality also allows full reproducibility in scientific papers, because the authors can pinpoint to a specific moment in OSM history and extract the data layers as a snapshot of given moment in time.

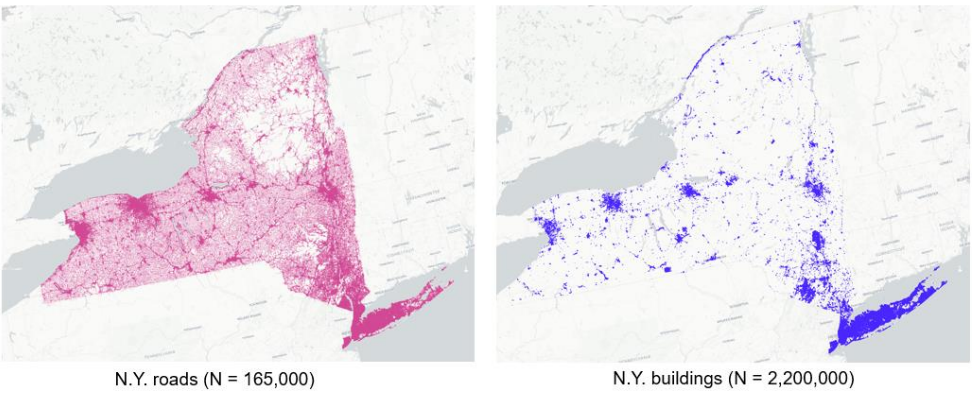


Fig. 1: Roads and buildings extracted from OSM with pyrosm -library covering the New York State.

Pyrosm is a useful tool to gather relevant geospatial data e.g. about built environment and services, which can be used as relevant input data for "Rapid Realistic Routing with R5 in Python" package called r5py. R5py is a Python wrapper for the Java-based R5 routing analysis engine (Conway et al., 2018) that allows realistic routing on multimodal transport networks (considering walk, bike, public transport and car). The tool is inspired by r5r (Pereira et al., 2021), a similar tool for R programming language, and the packages are developed in collaboration. R5py is designed to interact with geopandas GeoDataFrames, and it can be used together with pyrosm. R5py represents a simple but efficient way to run R5 routing engine which is originally developed by Conveyal LLC. Conveyal is the same company that has heavily contributed to the development of OpenTripPlanner. R5py uses OpenStreetMap and GTFS as input and constructs a routable multimodal graph based on these datasets. R5py can be used to generate detailed routing analyses or calculate travel time matrices using parallel computing, which are relevant information for studying various geographical phenomena, where spatial accessibility plays a role. The tool allows taking urban dynamics into consideration and the user can easily specify the date and time of day for the analysis. Hence, also historical analyses are possible with r5py (in a similar manner as with pyrosm), and you can e.g. investigate longitudinal changes in urban accessibility due to changes in transport infrastructures and schedules. R5py integrates seamlessly with modern data analysis workflows based on open source Python libraries having a linkage to geopandas.

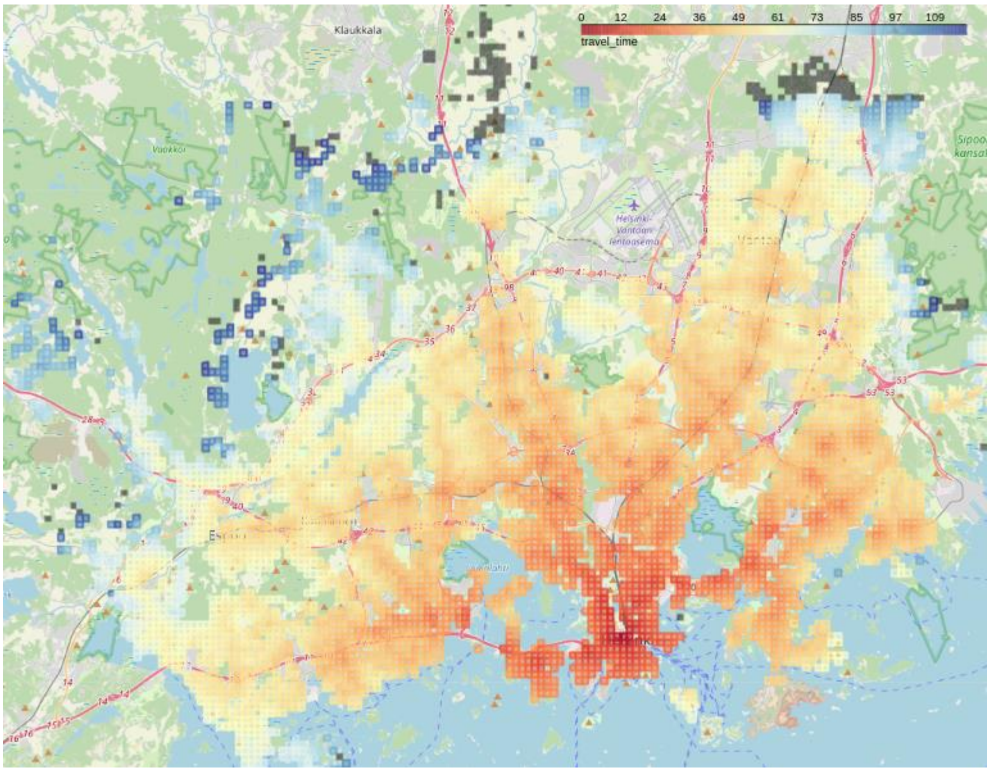


Fig. 2: Travel times to Helsinki city centre by public transport calculated with r5py -library.

In combination, pyrosm and r5py provide useful scientific software for urban researchers and practitioners alike. The libraries and their documentation are available at:

- https://pyrosm.readthedocs.io/en/latest/
- https://r5py.readthedocs.io/en/latest/

References

CONWAY, M.W., BYRD, A., EGGERMOND, M. VAN, 2018. Accounting for uncertainty and variation in accessibility metrics for public transport sketch planning. Journal of Transport and Land Use 11. https://doi.org/10.5198/jtlu.2018.1074

HAKLAY, M., WEBER, P., 2008. OpenStreetMap: User-Generated Street Maps. IEEE Pervasive Computing 7, 12–18. https://doi.org/10.1109/MPRV.2008.80

MAHAJAN, V., KUEHNEL, N., INTZEVIDOU, A., CANTELMO, G., MOECKEL, R., ANTONIOU, C., 2021. Data to the people: a review of public and proprietary data for transport models. Transport Reviews 0, 1–26. https://doi.org/10.1080/01441647.2021.1977414

PEREIRA, R.H.M., SARAIVA, M., HERSZENHUT, D., BRAGA, C.K.V., CONWAY, M.W., 2021. r5r: Rapid Realistic Routing on Multimodal Transport Networks with R5 in R. Findings. https://doi.org/10.32866/001c.21262

TENKANEN, H., 2021. HTenkanen/pyrosm: v0.6.1. Zenodo. https://doi.org/10.5281/zenodo.5561232

Generating indoor route instructions with multiple levels of detail

ZHIYONG ZHOU[1], ROBERT WEIBEL[1], KAI-FLORIAN RICHTER[2], STEPHAN WINTER[3] AND HAOSHENG HUANG[4]

1 Department of Geography • University of Zurich • Winterthurerstrasse 190 • 8057 Zurich, Switzerland

2 Department of Computing Science • Umeå University • 901 87 Umeå, Sweden

3 Department of Infrastructure Engineering • The University of Melbourne • Victoria 3010, Australia

4 Department of Geography • Ghent University • Krijgslaan 281, S8 • 9000 Gent, Belgium

Tel.: +41 (0)44 63 55229 • E-Mail: zhiyong.zhou@geo.uzh.ch

Keywords: Indoor route instruction, levels of detail, hierarchical navigation guidance

Summary: *Existing mobile indoor navigation systems primarily provide human wayfinders turn-by-turn route instructions on a uniform level of detail. In realistic navigation guidance, however, wayfinders may often prefer more concise route instructions, especially if they are (particularly) familiar with a building. While it would be useful to develop an indoor navigation system that can provide users with multiple levels of detail of indoor route instructions, there are few approaches to generating such a fashion of indoor route instructions. In this extended abstract, we propose an approach to generating indoor route instructions with multiple levels of detail. The approach is composed of three steps: route matching to a given hierarchical indoor data model, textualization using proposed schemas, and spatial relation inference for the textualization. The approach enables the automatic generation of indoor route instructions for hierarchical indoor navigation guidance and thus improves the usability of indoor navigation systems.*

Introduction

People commonly communicate route information at multiple levels of detail (LoDs) rather than using highly detailed turn-by-turn directions (Tenbrink and Winter, 2009). The advantage of route communication at multiple LoDs is that it results in more concise route instructions that adapt to route receivers' or wayfinders' background knowledge and thus facilitate the understanding and memorizing of communicated route information as well as improve navigation performance (Daniel and Denis, 2004; Lovelace et al., 1999). This also applies to indoor environments in addition to outdoor environments (Winter et al., 2018).

However, existing indoor navigation systems, such as MazeMap (https://use.mazemap.com/), primarily provide users with turn-by-turn route directions. Although some attempts have been to enrich semantic information of route directions by including indoor landmarks (e.g., "turn right after the elevator, you will pass through one door"; see Fellner et al., 2017), the instructed routes are still restricted to each segment. Therefore, it is likely to offer excessive route information to users and increase their workload to memorize and follow route instructions.

By contrast, indoor route instructions with multiple LoDs are able to abstract the information of route segments to other indoor spatial elements (e.g., floor, corridor, corridor part, intersection, side) that are understandable to humans and benefit users to choose the appropriate level of route instructions. For example, an instruction such as "go to the 3rd floor " may be enough for staff working in the building. However, if a user who visits the building for the first time does not know the locations of elevators or staircases, more detailed information to find an elevator in the building, such as "turn right at the third section of the corridor, you will find an elevator ", can be further provided.

We have developed a hierarchical indoor data model to support the generation of indoor route instructions with multiple LoDs (Zhou et al., 2022). However, computational methods are still missing regarding how various levels of elements in the hierarchical model can be translated into appropriate texts of hierarchical indoor route instructions. Therefore, we propose a computational method to automatically generate indoor route instructions with multiple LoDs based on the given hierarchical data model.

Proposed approach

The approach is mainly composed of three steps: route matching, schema-based textualization, and spatial relationship inference. A given route (e.g., the shortest path between an origin and a destination) in a building is taken as input and matched to the adopted hierarchical indoor data model, resulting in a route extending over six levels of the hierarchy (except the building level).

In the second step, two types of route instruction schemas, destination descriptions, and route directions are used to convert the information of route points, the inferred directions, and ordering relationships into textual route instructions. The former paradigm focuses on the “where” question in route communication, while the latter paradigm answers the “how” question in route communication.

Finally, we compute the spatial relations between segments of the matched route. In this work, we mainly consider two types of spatial relations for indoor route instructions: egocentric directions including left, right, up, and down, and ordinal relations such as first, and second. The spatial relation inference algorithms are thus developed to identify the specific spatial relation terms required for textualization.

Case study

We conducted a case study on the Y25 building at the University of Zurich to demonstrate the proposed approach for the generation of indoor route instructions with multiple levels of detail. The test route is shown in Fig. 1 and the corresponding generated route instructions with multiple LoDs are given in Table 1. Since the work is still ongoing, we expect to present the comparison results between the generated route instructions and typically used flat turn-by-turn route directions to evaluate the benefits and limitations of our implemented results at the conference.

Table 1. The generated route instructions at six levels of detail for test route

Levels	Results
Level 1 (floor)	1. Start from the main entrance 2. Take the staircase to L floor 3. Coffee room is on that floor
Level 2 (axial)	1. Start from the main entrance 2. Take the staircase to L floor 3.1 Exit the staircase 3.2 Turn to the right corridor 3.3 Coffee room is along that corridor
Level 3 (segment)	1. Start from the main entrance 2. Take the staircase to L floor 3.1 Exit the staircase 3.2 Turn to the right corridor 3.3.1 Go along the corridor to the third part of the corridor 3.3.2 Coffee room is situated in that part
Level 4 (junction)	1. Start from the main entrance 2. Take the staircase to L floor 3.1 Exit the staircase 3.2 Turn to the right corridor 3.3.1 Go along the corridor to the third part of the corridor 3.3.1.1 You will pass two fire emergency doors 3.3.2 Coffee room is situated in that part
Level 5 (side)	1. Start from the main entrance 2. Take the staircase to L floor 3.1 Exit the staircase 3.2 Turn to the right corridor

	3.3.1 Go along the corridor to the third part of the corridor 3.3.1.1 You will pass two fire emergency doors 3.3.2 Coffee room is situated in that part 3.3.2.1.1 You will find it on your left side
Level 6 (basic)	1. Start from the main entrance 2. Take the staircase to L floor 3.1 Exit the staircase 3.2 Turn to the right corridor 3.3.1 Go along the corridor to the third part of the corridor 3.3.1.1 You will pass two fire emergency doors 3.3.2 Coffee room is situated in that part 3.3.2.1.1 You will find it on your left side 3.3.2.1.1.1 It is the first room

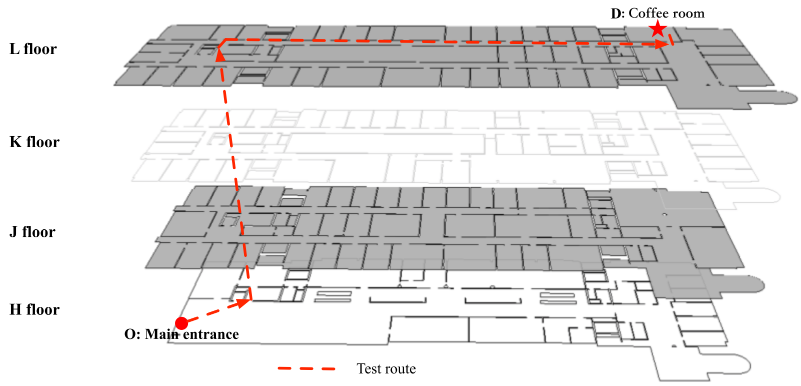


Fig. 1: Test route of the case study

Conclusions

This extended abstract presents an automatic approach to generating indoor route instructions with multiple levels of detail, which addresses the limitation of existing turn-by-turn fashion in indoor navigation systems and allows users to choose the approximate content of navigation guidance in buildings. In the future, we will investigate the relationships between a certain LoD of indoor route instructions and the familiarity of users.

References

DANIEL, M. P., & DENIS, M., 2004: The production of route directions: Investigating conditions that favour conciseness in spatial discourse. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 18(1), 57-75.

FELLNER, I., HUANG, H., & GARTNER, G. 2017: “Turn left after the WC, and use the lift to go to the 2nd floor”—Generation of landmark-based route instructions for indoor navigation. ISPRS International Journal of Geo-Information, 6(6), 183.

LOVELACE, K.L., HEGARTY, M. & MONTELLO, D.R., 1999, August: Elements of good route directions in familiar and unfamiliar environments. In International conference on spatial information theory (pp. 65-82). Springer, Berlin, Heidelberg.

TENBRINK, T., & WINTER, S., 2009: Variable granularity in route directions. Spatial Cognition & Computation, 9(1), 64-93.

WINTER, S., HAMZEI, E., WEGHE, N. V. D., & OOMS, K. 2018: A graph representation for verbal indoor route descriptions. In German Conference on Spatial Cognition (pp. 77-91). Springer, Cham.

ZHOU, Z., WEIBEL, R., RICHTER, K. F., & HUANG, H., 2022: HiVG: A hierarchical indoor visibility-based graph for navigation guidance in multi-storey buildings. Computers, Environment and Urban Systems, 93, 101751.

Building Virtual Knowledge Graphs from CityGML Data

LINFANG DING[1] AND GUOHUI XIAO[2] AND HONGCHAO FAN[1] AND DIEGO CALVANESE[3,4] AND LIQIU MENG[5]

[1] Norwegian University of Science and Technology • 7491 Trondheim • Norway

[2] University of Bergen • 5007 Bergen • Norway

[3] Free University of Bozen-Bolzano • 39100 Bolzano • Italy

[4] Umeå University • 90187 Umeå • Sweden

[5] Technical University of Munich • 80333 Munich • Germany

E-Mail: linfang.ding@ntnu.no

Keywords: CityGML, Ontology, Virtual Knowledge Graph

Summary: *In this work, we show how to expose CityGML data as a Virtual Knowledge Graph (VKG). We use 3DCityDB to store the CityGML data, and Ontop to build the VKG. We demonstrate the workflow using the models of the main campus of Technical University of Munich.*

Introduction

3D city models have been increasingly employed for advanced visualization and analysis tasks in various LBS applications, including indoor navigation and emergency rescue (Sun et al., 2020), and virtual and augmented reality (Santana et al., 2017). A widely adopted standard for the representation and exchange of 3D city models is *CityGML* (City Geography Markup Language) by Open Geospatial Consortium (OGC). It defines the three-dimensional geometry, topology, semantics, and appearance of the most relevant topographic objects in urban or regional contexts. Specifically, the representation of semantic and topological properties distinguishes CityGML from pure graphical 3D city models and enables thematic and topological queries and analyses.

One objective of CityGML is to inter-relate the 3D city information with other data to create a more complete representation of the urban landscape (Kutzner et al., 2020). However, this has not been exploited much in the research community. In this work, we tackle this problem by using semantic web technologies, and move urban data into *Knowledge Graphs (KGs)*. Thanks to the flexibility of the graph structure of KGs, multiple KGs can be easily integrated when they use vocabularies shared through ontologies. Many studies proposed geo-ontologies and so-called *GeoKGs* to represent domain knowledge and support geospatial data integration. Most of these works integrate the geodata sources by converting and materializing the original data as an RDF graph, and then storing such graph in an RDF store (Vinasco-Alvarez et al., 2020). Due to the resulting duplication of data, this way of proceeding can be expensive, especially when data sets are large or change frequently.

To overcome the challenges posed by materialization of the RDF graph, a different approach has been proposed, called *Virtual Knowledge Graph (VKG)* (Xiao et al., 2019). VKG is a popular paradigm that enables end users to access data sources through an ontology, which is semantically linked to the data sources by means of a *mapping*. Such mapping is expressed in the R2RML language standardized by the W3C. Thus, the ontology and mapping together, called a *VKG Specification*, expose the underlying data source as a virtual RDF graph, and make it accessible at query time using the standard W3C query language SPARQL. Using knowledge representation and automated reasoning techniques, a VKG system will then reason about the ontology and mapping and reformulate the SPARQL queries in terms of queries that can be directly evaluated at the data sources. This makes it possible to avoid the high cost of materialization. Such VKGs can be used in several application areas for LBS, such as dynamic semantic integration of urban information, and smart queries for PoI recommendations in routing algorithms.

Methodology

We proposed a framework called *CityGML VKG* as illustrated in Fig.1. We store the CityGML data into a relational database, and create a VKG specification (i.e., an ontology and a mapping). Then we use the popular VKG system *Ontop* (Calvanese et al., 2017) to expose the CityGML data as a VKG, which can be queried using the standard SPARQL query langauge.

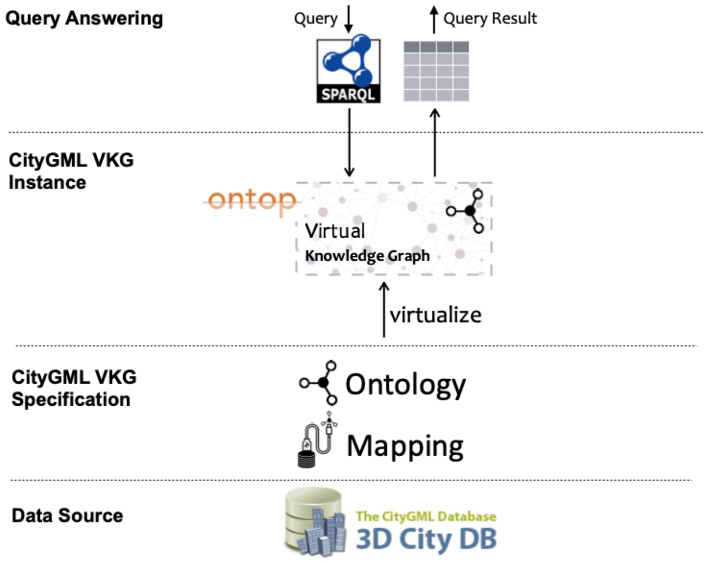


Fig. 1: Framework of CityGML VKG

Test data

We use the CityGML data of the main campus of Technical University of Munich as the test data (Fig 2). We make use of the 3DCityDB project, which implements the standard SQL encoding of CityGML, and import the sample data to 3DCityDB using PostgreSQL as a backend.

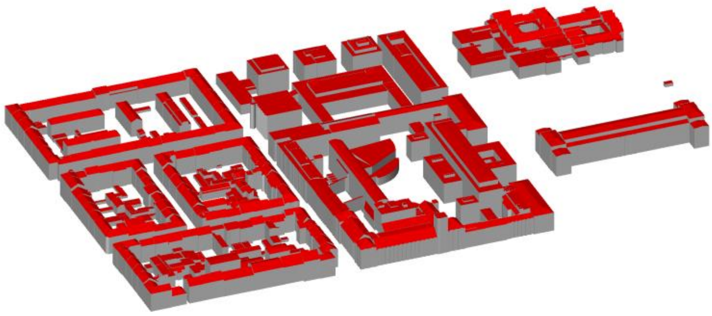


Fig. 2: The main campus of Technical University of Munich

Virtual Knowledge Graph Creation

We adopt the CityGML ontology[1] created by the University of Geneva, shown in the left part

[1] http://cui.unige.ch/isi/onto//citygml2.0.owl

of Fig. 3. We have developed a suitable R2RML mapping (the right part of Fig. 3) to 3DCityDB using the ontology editor Protégé with the Ontop plugin.

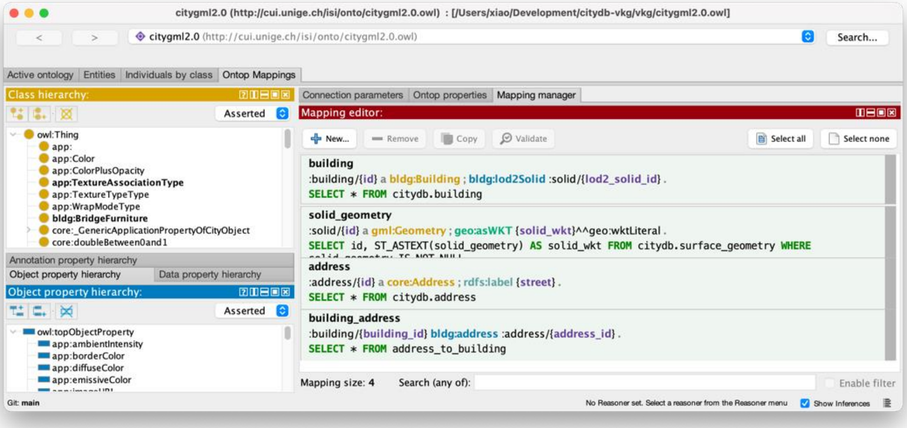


Fig. 3: The ontology and mapping editor in Protégé with the Ontop Plugin

Example Query

We show one example query in Fig 4, which retrieves all the buildings, together with their addresses and the LOD 2 solids. When evaluating this query, Ontop translates it into a SQL query, and sends it to the 3DCityDB backend.

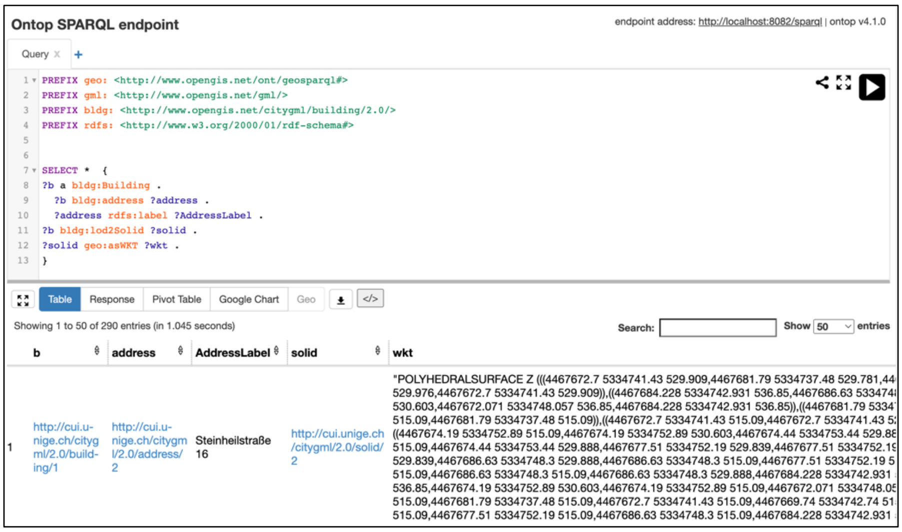


Fig. 4: An example SPARQL query evaluated in the SPARQL endpoint of Ontop

Conclusions

We have developed a proof-of-concept system for exposing CityGML data as a VKG. In the future, we will extend the coverage of the mapping, intregrate other datasets into the VKG, and apply the technology in LBS applications.

Acknowledgment

This work is supported by the D2G2 project jointly funded by DFG and the Province of South Tyrol, Italy.

References

CALVANESE, D., COGREL, B., KOMLA-EBRI, S., KONTCHAKOV, R., LANTI, D., REZK, M., RODRIGUEZ-MURO, M. & XIAO, G., 2017: Ontop: Answering SPARQL Queries over Relational Databases. – Semant. Web J., 8(3): 471–487, https://doi.org/10.3233/SW-160217.

KUTZNER, T., CHATURVERDI, K. & KOLBE, T. H., 2020: CityGML 3.0: New Functions Open Up New Applications. – PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 88: 43–61.

SANTANA, J.M., WENDEL, J., TRUJILLO, A., Suárez, J.P., SIMONS, A., KOCH, A., 2017: Multimodal Location Based Services – Semantic 3D City Data as Virtual and Augmented Reality. In: Gartner, G., Huang, H. (eds) Progress in Location-Based Services 2016. Lecture Notes in Geoinformation and Cartography. Springer, Cham. https://doi.org/10.1007/978-3-319-47289-8_17.

SUN, Q., Zhou, X. & HOU, D., 2020: A Simplified CityGML-Based 3D Indoor Space Model for Indoor Applications. *Appl. Sci.* *10*, 7218. https://doi.org/10.3390/app10207218.

VINASCO-ALVAREZ, D., SAMUEL, J., SERVIGNE, S., & GESQUIER Eesquière, G., 2020: From CityGML to OWL. – Technical Report LIRIS UMR 5205, hal-02948955

XIAO, G., DING, L., COGREL, B., & CALVANESE, Diego., 2019: Virtual Knowledge Graphs: An Overview of Systems and Use Cases. – Data Intelligence, 1:201–223.

DP Mobility Report: A Python Package for Mobility Data Explorations with Differential Privacy Guarantees

ALEXANDRA KAPP[1], SASKIA NUÑEZ VON VOGT[2], HELENA MIHALJEVIĆ[1] AND FLORIAN TSCHORSCH[2]

[1]Hochschule für Technik und Wirtschaft Berlin, University of Applied Sciences • Wilhelminenhofstr. 75A • 12459 Berlin

[2]Distributed Security Infrastructures • Technische Universität Berlin • Straße des 17. Juni 135 • 10623 Berlin

E-Mail: alexandra.kapp@htw-berlin.de

Keywords: human mobility data, differential privacy, user-level privacy, exploratory data analysis, mobility report

Summary: *Any data analysis commonly begins with data exploration which involves much repetitive work as most explorations include similar steps. Tools have been developed for tabular data to accelerate and facilitate those procedures (e.g., Brugman, 2019). While of great use for regular tabular data, these tools are rather poorly applicable to human mobility data as, e.g., the average of latitude or longitude typically does not provide relevant insight. Graser (2021) presents a protocol for an exploratory data analysis (EDA) for continuous movement data with the goal to identify problems, such as unrealistic jumps in GPS traces. Our work, in contrast, focuses on EDAs for origin-destination trips of urban human mobility data. We introduce a open-source Python package that provides such standardized explorations as a* Mobility Report*.*

Human mobility data becomes increasingly available from various sources, e.g., through the rise of location-based service apps, shared mobility, or the digitalization of public transport, for example, with the use of smart cards. This data is of great importance for an informed and efficient design of sustainable urban mobility (Creutzig, 2021). Thus, in addition to the use of initial explorations for data analysts, summary reports are increasingly released to third parties, such as municipal administrations or in the context of citizen participation, and serve as a basis for decision-making processes. However, it is often overlooked that such explorations already pose a threat to privacy as they reveal potentially sensitive location information, which is why even aggregated statistics should not be shared without further privacy measures (Nuñez von Voigt et al., 2020). We thus provide differential privacy guarantees for the reports produced with the proposed dp_mobility_report *package to share mobility statistics in accordance with state-of-the-art privacy measures.*

Within this work, we elaborate on the implementation of the code package. Two mandatory user inputs are needed to create a report: the raw data itself and an amount of privacy budget. Additional optional parameters can be used to customize the report. The produced report, provided as an HTML-file, is structured into the four segments overview, place analysis, origin-destination analysis, *and* user analysis *that comprise the most common analyses typically conducted with urban human mobility datasets. The overview provides basic information on record, trip, and user counts, in addition to temporal properties. The place analysis segment entails information about the spatial distribution of the data and the most visited locations. The origin-destination section informs about the most commonly traveled connections, travel times, and distances. The user analysis segment gives insights on statistics related to users, such as the radius of gyration, number of unique locations visited per user or the mobility entropy. We provide detailed information about each segment, including the respective statistics and visualizations.*

Furthermore, we give insights on different issues that must be considered with the implementation of differential privacy: Firstly, we discuss the issue of user-level privacy. Intuitively, the more records a user contributes to a dataset the more likely they will be identified. When differential privacy is secured by adding noise to the data, a high number of records per user requires a larger amount of noise and thus reduces the utility of the generated statistics (Amin et al., 2019). This poses the challenge of handling arbitrarily large user contribution as in general, an upper limit is not known upfront. We solve this issue by setting an upper bound and removing exceeding records. The number of maximum contributions per user can be set with an input parameter by the data analyst.

Secondly, leaking true minimum and maximum values, e.g., the minimum and maximum travel time, violates differential privacy. Thus, we use the exponential mechanism (McSherry & Talwar, 2007). to produce five-number summaries and cut related statistics, e.g., the travel time histogram, according to the differentially private minimum and maximum values. To provide differentially private histograms, Laplace noise is added to single histogram bins.

Thirdly, differential privacy makes use of a privacy budget which defines how much information can be shared about individuals within the dataset. The higher the budget, the less noise is added and thus the less the privacy is protected. If data is used for more than one analysis, the privacy budget needs to be split between all analyses to stay within budget. For a report which compiles a series of analyses, the question of an optimal budget splitting arises. We propose a default split that can be customized through user input. Additionally, the package offers the option to remove analyses from the report so that more budget can be spent on only selected analyses.

Lastly, sparse data produces especially noisy results. Implemented solutions for the communication of noise levels and uncertainty within the report are described and discussed in this work.

A workshop with mobility data practitioners provided a first evaluation of the package in terms of functionality and use cases.

References

AMIN, K., KULESZA, A., MEDINA, A. M., & VASSILVITSKII, S., 2019: Bounding User Con-tributions: A Bias-Variance Trade-off in Differential Privacy. Proceedings of the 36th International Conference on Machine Learning, 97, 263–271.

BRUGMAN, S., 2019: pandas-profiling: Exploratory data analysis for python. https://github.com/pandas-profiling/pandas-profiling

CREUTZIG, F., 2021: From smart city to digital urban commons: Institutional considerations for governing shared mobility data. Environmental Research: Infrastructure and Sustainability, 1(2), 025004.

GRASER, A., 2021: An exploratory data analysis protocol for identifying problems in continuous movement data. Journal of Location Based Services, 15(2), 89-117. doi:10.1080/17489725.2021.1900612.

MCSHERRY, F., & TALWAR, K., 2007: Mechanism Design via Differential Privacy. 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), 94–103. https://doi.org/10.1109/FOCS.2007.66

NUÑEZ VON VOIGT, S., PAULI, M., REICHERT, J., & TSCHORSCH, F., 2020: Every Query Counts: Analyzing the Privacy Loss of Exploratory Data Analyses. DPM '20: Data Privacy Management, Cryptocurrencies and Blockchain Technology - {ESOR-ICS} 2020 International Workshops, 12484, 258--266.

Predicting the transport pathway of dust storms using convolutional neural network

Mahdis yarmohamadi[1], Mohammad sharif[2,*], Ali asghar alesheikh[1] and farid karimipour[3]

[1]Department of Geospatial Information Systems • K. N. Toosi University of Technology • 15433-19967 Tehran • Iran

[2]Department of Geography • University of Hormozgan • 3995 Bandar Abbas • Iran

[3]Institute of Science and Technology (IST Austria) • 3400 Klosterneuburg • Austria

Tel.: +98 (0)76 3371 1000 • E-Mail: m.sharif@hormozgan.ac.ir

Keywords: Moving process, prediction, deep learning, CNN, dust storm

Summary: *In the dust storm (DS) process, analysing and predicting the movement of dust are crucial because it reveals their transport pathway and examines the next vulnerable areas to the dust event. By adopting a deep convolutional neural network (CNN) method, this study aims to predict the pathway of DSs that occur in an arid region in central and south Asia. 30 dust storms are extracted from the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) product to train the model and evaluate the results. The overall accuracy of 0.973 for the next 10 hours shows that CNN can establish an accurate prediction for the dust transport pathway.*

Introduction

Moving processes (e.g., dust storms and fire spread), in contrast to moving point objects (e.g., humans and vehicles), are areal phenomena whose sizes constantly changing. Robust methods are required for not only analyzing the transport pathway (trajectory) of moving processes but also for modeling their dynamic areas (Boroumand et al., 2022; Goudarzi et al., 2022). Among moving processes, dust storm (DS) is a common source of mineral aerosols and is recognized as a serious environmental hazard, especially in arid and semi-arid regions (Prospero et al., 2002). Dust movement prediction is an important part of controlling air quality, by which we can take steps towards decreasing the impact of dust on the atmosphere.

Owing to the remote sensing technology, monitoring DSs is possible in different spatial and temporal resolutions. Aerosol Optical Depth (AOD) is a basic parameter that reflects aerosol optical properties in satellite images. MERRA-2 is a reanalysis data product from NASA to observe the earth, and the MERRA-2 AOD observes aerosol at 0.5°×0.625° resolution since 1980 (MERRA, 2015). Therefore, it is a suitable product for hourly monitoring of DS movements. On the other hand, deep learning methods are becoming increasingly popular in DS movement prediction. Deep convolutional neural networks (CNNs) have been substantiated by other researchers as a model to achieve more accurate outcomes for DS analysis and pathway estimation (Jiao et al., 2021). Therefore, the objective of this research is to predict the transport pathway of DSs using MERRA-2 dataset and a deep CNN method.

Material and Method

An arid study area in Turkmenistan, Iran, Afghanistan, and Pakistan is chosen (Fig. 1). The occurrence of 30 DSs for a total of 1858 hours between 2000 and 2021 are derived from the media and validated with the MODIS AOD dataset, and the hourly MERRA-2 AODs are extracted. Because we are facing a time series problem, the data is divided into input and output categories. The input dataset contains images of MERRA-2 AOD with 140×143 pixels at t time step and the output dataset contains visually labelled images from t+1 to t+10 time steps. This means that the data of the last hour is used to predict the next 10 hours of the DS.

In the CNN architecture, the main task of the convolution layers is learning the features from input data. In this research, nine (3×3) convolutional layers consisting of 64, 128, 32, 32, 32, 32, 32, 32, and 16 filters are selected. The extracted features are flattened and passed into two dense layers consisting of 64 and 140×143 neurons, respectively. For all the layers, the Rectified Linear Units (ReLUs) are used as activation functions (Glorot & Bordes, 2011), except for the last layer that uses sigmoid. L2 regularization with value of 0.0001 and Adam optimizer with a batch size of 4 are used (Kingma & Ba, 2015). The learning rate of 0.0005 is considered during the training process and binary cross-entropy has been used as a cost function. Fig. 2 shows the three stages of model input, feature extraction, and model output.

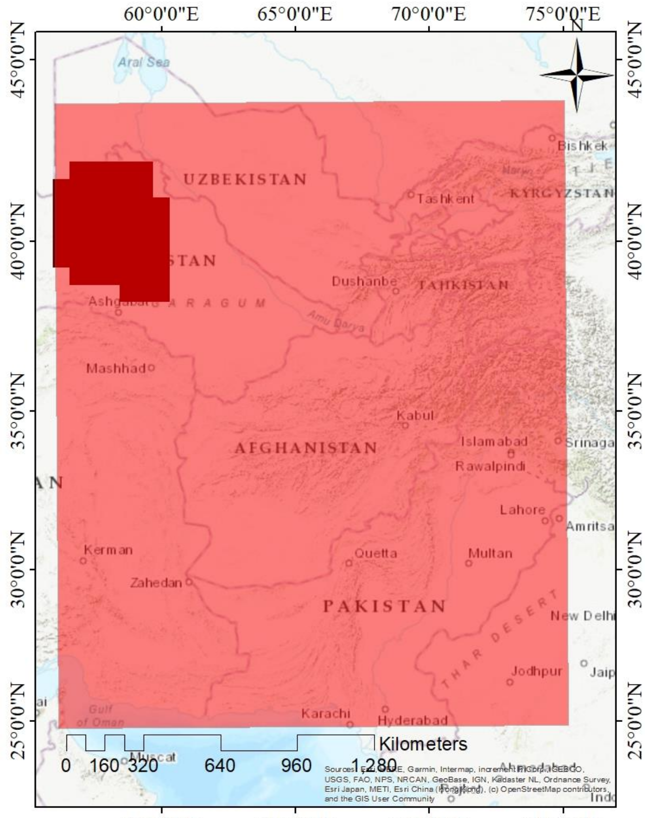


Fig. 1: Study area (in light red) and a sample DS that originates from Turkmenistan (in dark red)

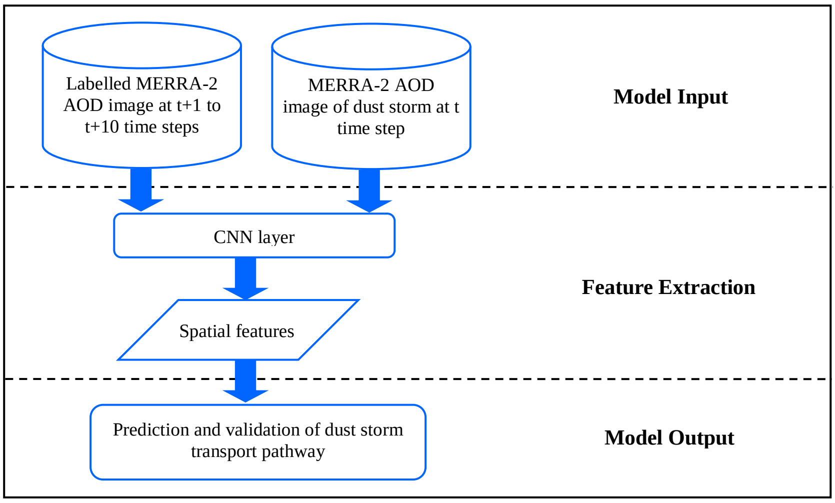


Fig. 2: Framework for DS transport pathway prediction

Results

90% and 10% of the data are used for training and validation of the model, respectively. Fig. 3 shows the sample confusion matrices of the 1st, 5th, and 10th time steps. The evaluation of the model is based on the conformity between prediction and reality that is measured by the Jaccard distance, which measures dissimilarity between sample sets. The overall accuracies of DS predictions of the next 10 time steps are shown in Table 1. The results justify the robustness of the CNN method in transport pathway prediction of DSs and examining the next vulnerable areas to the dust event.

Actual		Predicted: DS	Predicted: Not DS
	DS	94.47%	5.53%
	Not DS	2.08%	97.92%

(a)

Actual		Predicted: DS	Predicted: Not DS
	DS	97.05%	2.95%
	Not DS	1.99%	98.01%

(b)

Actual		Predicted: DS	Predicted: Not DS
	DS	96.39%	3.61%
	Not DS	2.07%	97.93%

(c)

Fig. 3: Confusion matrices of three time steps, (a) t+1, (b) t+5, and (c) t+10

Table 1: Overall accuracy of DS predictions of the next 10 time steps

	1h	**2h**	**3h**	**4h**	**5h**	**6h**	**7h**	**8h**	**9h**	**10h**
Accuracy	0.9777	0.9791	0.9796	0.9796	0.9796	0.9793	0.9792	0.9786	0.9783	0.9773

References

MERRA, G., 2015: tavgU_2d_lnd_Nx: 2d, diurnal, time-averaged, single-level, assimilation, land surface diagnostics V5. 12.4. EarthData GES DISC NASA.

GLOROT, X. & BORDES, A., 2011: Deep Sparse Rectifier Neural Networks. – In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, PMLR 15: 315-323.

JIAO, P., WANG, J., CHEN, X., RUAN, J., YE, X. & ALAVI, A. H., 2021: Next-Generation Remote Sensing and Prediction of Sand and Dust Storms, State-of-the-Art and Future Trends. – International Journal of Remote Sensing 42 (14): 5281–5320.

KINGMA, D. P. & BA, J. L., 2015: Adam: A Method for Stochastic Optimization. – ArXiv Preprint ArXiv:1412.6980, 1–15.

PROSPERO, J. M., GINOUX, P., TORRES, O., NICHOLSON S. E. & GILL T. E., 2002: Environmental Characterization of Global Sources of Atmospheric Soil Dust Identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) Absorbing Aerosol Product. – Reviews of Geophysics 40 (1): 2-1-2–31.

BOROUMAND, F., ALESHEIKH, A. A., SHARIF, M. & FARNAGHI, M., 2022: FLCSS: A Fuzzy-Based Longest Common Subsequence Method for Uncertainty Management in Trajectory Similarity Measures. – Transactions in GIS 1–15.

GOUDARZI, S., SHARIF, M. & KARIMIPOUR, F., 2022: A Context-Aware Dimension Reduction Framework for Trajectory and Health Signal Analyses. – Journal of Ambient Intelligence and Humanized Computing 13 (5): 2621-2635.

Digital learning using LBS: the “CartoWalk” mobile application concept

JULIA EIGNER AND OLESIA IGNATEVA

Research Unit Cartography• Technische Universität Wien • Erzherzog-Johann-Platz 1/120-6 * • A-1040 Wien • E-Mail: e11804648@student.tuwien.ac.at

Keywords: mobile application, digital learning, digitalization, location based service

Summary: *This work proposes the concept of the mobile application for learning cartography. By using location-based services, it is possible to playfully gain or strengthen cartographic knowledge while moving between different stations.*

Digital learning has become more topical because the pandemic has challenged the traditional learning approach. This project aims to bring the cartography class outdoors and learn cartography as it can be experienced outdoors with digital learning. The main idea of the mobile application "CartoWalk" is to explore certain places in Vienna city and to learn cartography. In addition, users can test existing knowledge and thus strengthen it.

A target group is a group of students who attend Cartography class. They have to navigate sights using the interactive map and then answer questions and solve the tasks about specific cartographic topics. These sights are called stations, which will be well-known landmarks in Vienna. Whenever the user moves, the path will be marked in the app, and it will always be visible where to find the next station.

Individual stations could cover different topics, for example, cartographic generalization. Example questions of this station could be what the users already know about generalization, why generalization is made, and how to recognize that a map has been well generalized. The users will also receive the questions related to a given map where they should identify which cartographic generalization operators have been used.

The topic of another station could be symbolization and graphical variables. Some questions could be why cartographers use map symbols and graphic variables. Each group will receive a different map, which will address specific topics, such as traffic, gastronomy or green spaces. These maps use symbols consisting of graphic variables that users can examine.

At the station about mapmaking, users can learn about map production and the principles used when creating the map. The station about map labeling will include a sample task to identify which map was labeled correctly based on different maps. At the station dedicated to the map’s layout, the target group should learn its elements and answer what map layout components were used.

With the help of all these stations, it should be possible to acquire basic cartographic knowledge by using the map. The application will teach users to deconstruct maps and be critical of the map design. Location-based services in mobile applications enable to find all stations easily and thus achieve an optimal learning effect.

The application will be developed by the end of September 2022 and will result in a bachelor’s thesis.

Method Development for the Visualisation of Bicycle Trajectories and Traffic Related Parameters by a Space-Time Cube

SYLVIA LUDWIG[1], ANDREAS KELER[2] AND CHENYU ZUO[3]

[1]Technical University Munich • E-Mail: sylvia.ludwig@tum.de

[2]Chair of Traffic Engineering and Control • Technical University Munich • E-Mail: andreas.keler@tum.de

[3]Chair of Cartography and Visual Analytics • Technical University Munich • E-Mail: chenyu.zuo@tum.de

Keywords: data processing, data visualisation, space-time cube, traffic analysis, bicycle trajectories

Abstract

To establish the bicycle traffic as a climate friendly alternative to the motorised private transport, a rethinking of city planning in terms of traffic is necessary. The analysis of existing spatiotemporal data from road users is crucial for planning streets and junctions in a way, which fits the needs of cyclists.

In the present thesis spatiotemporal data of cyclists at the junction of Theresienstraße and Ludwigstraße in Munich is visualised by bicycle trajectories in a space-time cube (STC). The data, which is used for the visualisation, is collected by a video camera with a frame rate of 30 frames per second. After converting each individual video frame to an orthophoto, road users are detected using computer vision methods, localised and tracked between the georegistered video sequences (Adamec et al., 2018). For the upcoming visualisation of the trajectories merely the 260 bicycles of a total of 1113 road users are considered, which corresponds to one half of the recorded cars in the selected time frame of 17 minutes and 40 seconds. The raw dataset is reduced and pre-processed into a transformed dataset of reduced memory storage to improve the performance of the further visualisation methods.

To account for a user-defined filtering function and other visualisation options, a graphical user interface (GUI) is implemented in MATLAB. By using this tool, the dataset of the aforementioned crossing can better be analysed in terms of the cycling behaviour. The GUI contains different tools, such as a feature, which filters the visualised trajectories for the origin and destination of the bicycles, therefore it is possible to analyse a specific bicycle stream. Another filtering feature, which is realised by a double slider, allows for the reduction of the amount of shown trajectories. Moreover, one can display the traffic lights, which are switched for the respective traffic flow. The traffic light circuit is visualised by spherical dots in green or red, which are shown as long as a cyclist sees the traffic light, see Figure 1. Two other features are the visualisation of the direction of traffic through small arrows and the possibility to show the velocity of the road users at a specific position, through a colouring of the trajectories. With the help of a colourbar, the respective speed from the colour of the trajectory can be read off.

By using the features of the GUI it possible to get a clearer picture of concrete driving patterns, which are visualised within the STC. Especially, the possibility of a data reduction by the aforementioned filtering features is crucial. It allows the user to keep track of the huge number of trajectories and analyse the behaviour of single cyclists. One specific pattern of the observed road users is the disregard of the red traffic light, which can be seen in Figure 1. The cyclists of the straight traffic flow of the Ludwigstraße, which are ignoring the red light are marked by 1 and 2. After a detailed analysis one can see, that these cyclists crossed the red traffic lights with a high velocity of around 20 km/h. A possible explanation for this behaviour can be determined with the help of the available video material. By using this, it becomes apparent, that the cyclists were not under immediate danger, as there were no motorised vehicles in the field of view. By a further analysis of the available dataset other

patterns can be observed with the implemented tool. One example is the driving behaviour at traffic lights, which can be best analysed by considering the bicycle velocities in the colourmap. This could be useful to help matching the traffic lights to cyclists and therefore to improve the traffic flow for this group of road users. One last pattern, which can be seen in the STC is the crossing behaviour of the cyclists, which is characterised by its diffuse picture. Part of the cyclists choose the possibility to use the same way as the motorised vehicles others the same way as the pedestrians. A special trend, which way is preferred by the cyclists cannot be ascertained.

With the visualisation method of the space-time cube it is possible to visualise traffic parameters such as the time to collision (TTC) and the post encroachment time (PET). In this work, the two parameters were determined using two different trajectory examples. The visualisation of the TTC and PET is quite challenging to automate and therefore difficult to visualise for all the trajectories at the same time.

Besides the analysis of the dataset itself, an expert interview was conducted to assess the usability of the GUI. Therefore, five contributors of the chair of traffic engineering and one contributor of the chair of cartography at Technical University Munich were interrogated, after they had a chance try out the features of the GUI. The experts' suggested improvements can be included for further work.

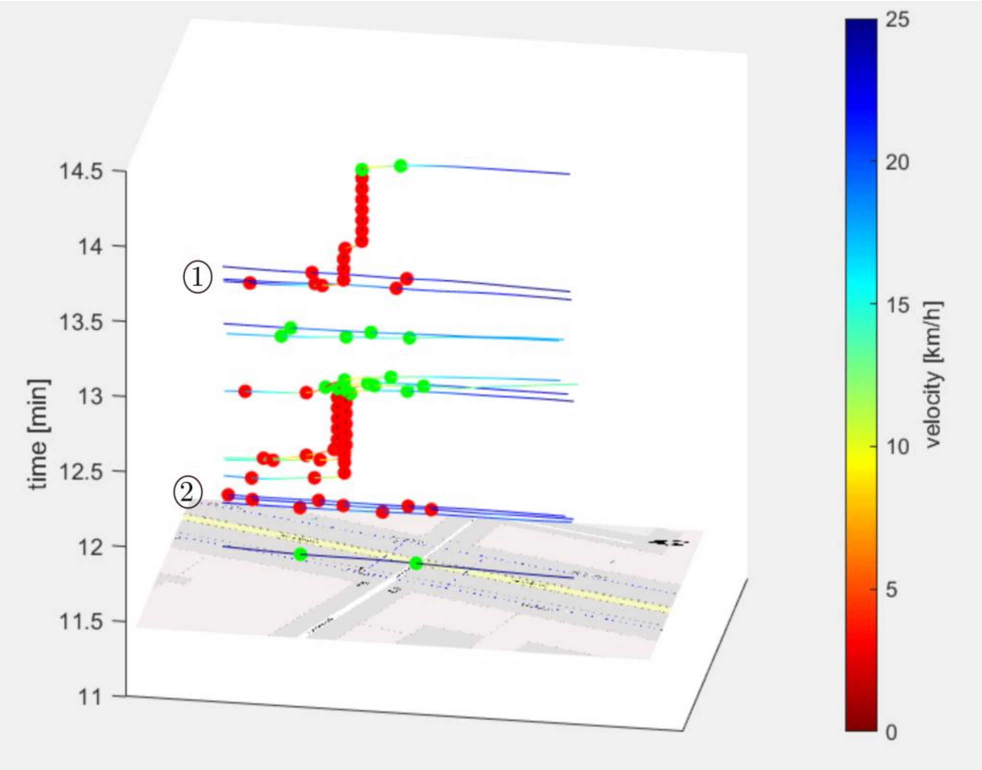


Fig. 1: Disregard of a red traffic light of the straight traffic flow of the Ludwigstraße

References

ADAMEC, V., SCHULLEROVA, B., HERMAN, D. & VEMOLA, A., 2018: Possibilities of the traffic emission monitoring in context of the smart cities concept. University of Pardubice, ISBN: 978-80-7560-152-0. URL: https://dk.upce.cz/handle/10195/71821.

Preliminary study of indoor emergency path planning based on fire emergency knowledge graph

JINGYI ZHOU, JIE SHEN*, LITAO ZHU, XUEWEI YAN AND WEIHONG SUN

School of Geography • Nanjing Normal University • 210023 Nanjing

Tel.: +86-025-85891347 • E-Mail: shenjie@njnu.edu.cn

Keywords: Emergency path planning, fire emergency, knowledge graph

Summary: *Fire is one of the frequent emergencies that threaten people's lives. The knowledge graph is an important tool to reveal entities and their interrelationships, and it plays an increasingly important role in emergency disaster reduction. At present, there are few knowledge graphs constructed in the field of fire emergency. This paper preliminarily explored the construction method of fire emergency knowledge graph, and the key technologies mainly include knowledge extraction, information fusion, and knowledge storage. Finally, the application direction of fire knowledge graph in indoor emergency path planning is discussed, and the three-dimensional fire simulation is realized by taking an office building as an example.*

Introduction

Fire is one of the disasters that occur most frequently and threat to human life. The high temperature and toxic smoke produced by fire combustion will seriously affect the physiology and psychology of the people, make the people lose their way in the building, and even cause casualties (Aleksandrov et al., 2018; Riboulet 2018). The main factors affecting personnel emergency include toxic gases, visibility, emergency system, etc. (Jeon et al., 2011; Butler et al., 2017).

The essence of knowledge graph is a semantic network that reveals the relationship between entities (Nickel et al., 2015). Usually, nodes are used to represent entities, concepts and events, and edges are used to represent relationships. When a disaster occurs, the data is effectively aggregated, fused, and stored based on the knowledge graph, which helps to quickly obtain useful information.

Disaster emergency rescue is usually participated by multiple social subjects. The knowledge graph can establish the network relationships among different fields, different entities, disaster data, and disaster events.

Construction of fire emergency knowledge graph

The knowledge graph construction process is shown in Figure 1. First, multi-source heterogeneous data is processed, and then knowledge such as entities, attributes, and relationships are extracted according to application scene, and various kinds of knowledge are classified and integrated. Finally, the knowledge graph is stored in the form of a graph structure.

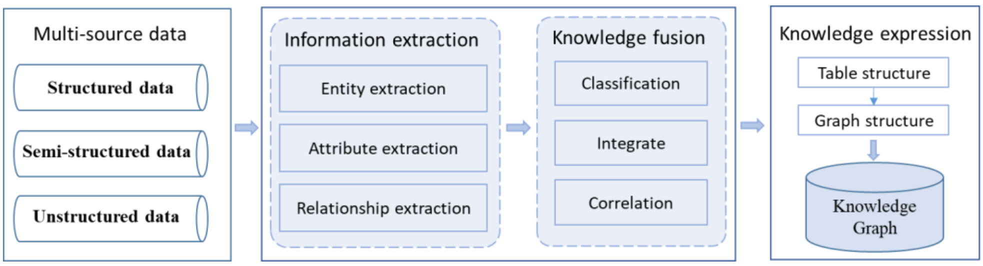


Fig. 1: Construction process of knowledge graph

This research focused on fire emergency, adopted the top-down manual modeling method to build a knowledge graph, and extracted the knowledge of entities, elements, and relationships for the models, data, combustion elements, emergency personnel involved in the fire. Figure 2 shows a part of the fire emergency knowledge graph.

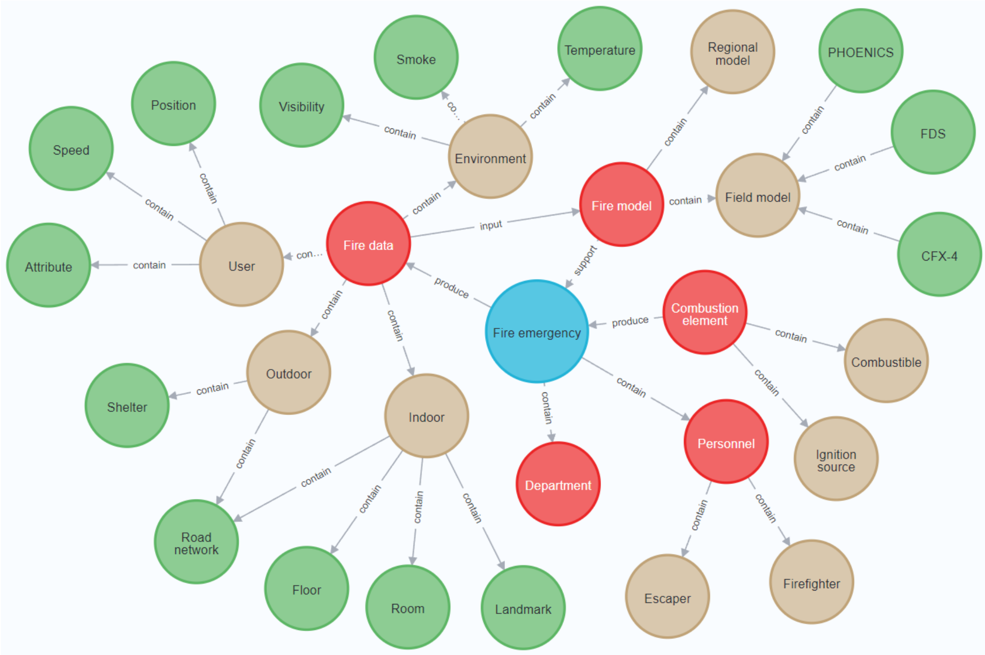


Fig. 2: Fire emergency knowledge graph (partial display)

Key idea of indoor emergency path planning

By analyzing the fire model, data, emergency departments, and other elements, after building the knowledge graph, we concluded that it has at least four applications:

(1) Knowledge fusion of fire numerical model and building information model

(2) Quickly realize the fusion of multi-source heterogeneous data

(3) User-oriented emergency path planning guidance

(4) Integration the emergency path and the fire emergency scene

According to the guidance of the constructed fire emergency knowledge graph, we have a research idea for indoor emergency path planning in fire scene, as shown in Figure 3:

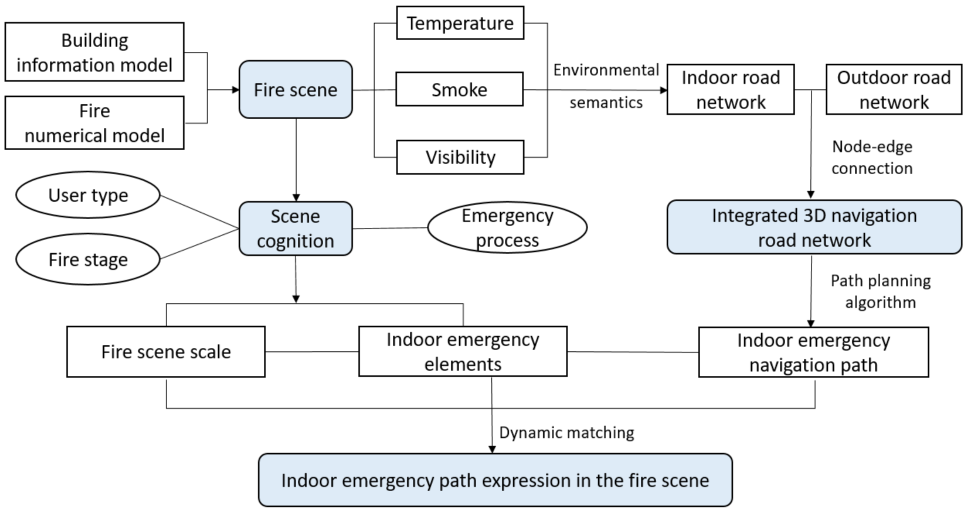


Fig. 3: Research ideas of indoor emergency path in fire scene

We choose an office building as the study area. According to the guidance of the constructed knowledge graph, the fire model can choose the FDS model and the fire environment data should at least include temperature, smoke, and visibility. Through

Pyrosim established a fire simulation scene, recording and analyzing visibility, temperature, and gas concentration in the fire smoke, the simulation results are shown in Figure 4.

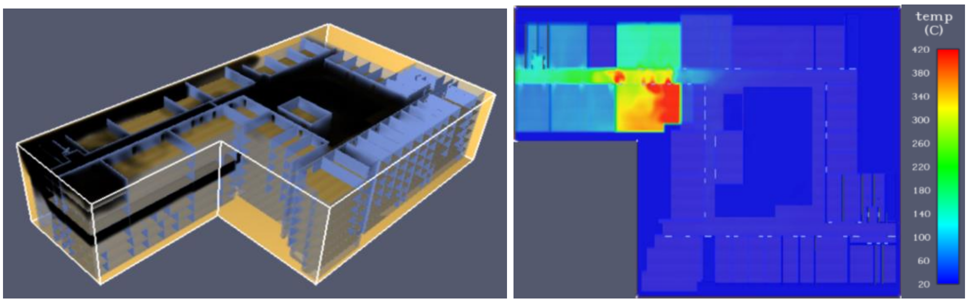


Fig. 4: Smoke diffusion and temperature distribution in the study area

Conclusion and future plan

This study has preliminarily constructed the fire emergency knowledge graph, and studied in the application direction of indoor emergency path planing. However, it is also necessary to conduct in-depth analysis of fire models, fire data, emergency management and other elemens to further improve the fire emergency knowledge graph. In the future, the knowledge graph will be applied to fire data fusion and emergency path planning to provide personal location-based emergency services in fire scene.

Acknowledgment

This research was funded by National Key R&D Program of China (2021YFE0112300), National Natural Science Foundation of China (NSFC) (No. 41871371), the Postgraduate Research and Practice Innovation Program of Jiangsu Province (No. KYCX22_1571).

References

ALEKSANDROV, M., RAJABIFARD, A., KALANTARI, M., LOVREGLIO, R., & GONZáLEZ, V. A., 2018: People choice modelling for evacuation of tall buildings. Fire technology, 54(5), 1171-1193.

BUTLER, K., KULIGOWKSI, E., FURMAN, S., & PEACOCK, R., 2017: Perspectives of occupants with mobility impairments on evacuation methods for use during fire emergencies. Fire Safety Journal, 91, 955-963.

JEON, G. Y., KIM, J. Y., HONG, W. H., & AUGENBROE, G., 2011: Evacuation performance of individuals in different visibility conditions. Building and Environment, 46(5), 1094-1103.

NICKEL, M., MURPHY, K., TRESP, V., & GABRILOVICH, E., 2015: A review of relational machine learning for knowledge graphs. Proceedings of the IEEE, 104(1), 11-33.

RIBOULET, T., 2018: Developing numerical model for Vierendeel failures of composite cellular beam in fire.

Design of classification early warning algorithm for navigation route waterlogging based on Doppler weather radar estimation of rainfall

SHUAI HONG[1,2,3,*], HONG PAN[4] AND JIJUN YANG[1,2,3]

1. Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China; hongsh_yc@njnu.edu.cn, Chloeyangxx@163.com

2. Jiangsu Center for Collaborative Innovation in Geographical Infor-mation Resource Development and Application, Nanjing 210023, China

3. School of Geography, Nanjing Normal University, Nanjing 210023, China

4. Sanjiang University, Nanjing 210012, China; pan_hong@sju.edu.cn

Keywords: Navigation route waterlogging classification, Early warning algorithm,

Urban waterlogging, SWMM, Doppler weather radar

Summary: *In urban waterlogging disaster, the severity of road waterlogging will have a direct impact on the safety of urban residents' driving (Yin, 2016). However, the commonly used navigation software lacks the function of forecasting and early warning for route waterlogging in navigation paths, and cannot meet the need for users' safe driving in extreme weather conditions, such as heavy rain. In view of the above problems, this study intends to design a waterlogging classification early warning algorithm for the navigation route. Among them, the calculation and extraction of navigation route waterlogging depth information and its fusion application with navigation algorithm are the key problems of this research.*

Our research mainly includes the following four aspects:

(1) Quantitative estimation of rainfall data. Based on rainfall echo rates from Doppler weather radar data, we choose the Z-R relationship (where Z is the rain radar reflectivity and R is the rainfall intensity) to estimate rainfall in the study area (Sokol, 2021), and combined with the rain gauge data of the ground observation station to correct the rainfall data estimated by Doppler weather radar.

(2) Simulation of inundation in the study area. Based on the pipe network data, we divide the study area into multiple sub-catchments, and the model parameters are determined according to the recommended value range of the hydrological model manual based on the analysis of the actual situation such as the characteristics of the study area. Then, combining the SWMM hydrological model with the surface DEM of the study area, and based on the terrain inundation algorithm, we construct an urban surface inundation module to simulate the urban waterlogging inundation process in the study area.

(3) Navigation route waterlogging depth calculation. By superimposing the distribution grid map of urban waterlogging in the study area with the distribution map of the urban road network, we can obtain the grid map of road water accumulation in the study area. Each cell value of the raster map represents the average water depth based on the average spatial elevation of the cell. Since water is a fluid, it can remain level everywhere as the water surface flows automatically, so we only need to know the actual elevation of the sampling point, and then offset it from the average elevation. Finally, based on the waterlogging classification standard for the vulnerability of the disaster-affected body (Li, 2015), we designed a road traffic classification rule that considers the depth of water accumulation.

(4) Design and Verification of Navigation Warning Algorithm. The user's location is constantly changing during the navigation process. This study uses the A^* algorithm to combine the user's location information to plan the user's navigation route in real time. In addition, the maximum water depth information max(depth) of the route sub-section is introduced into the A^* algorithm. Before executing the heuristic function f(n) of the A^* algorithm, we add the judgment of the road water depth passage rule. Based on this, we propose the navigation route waterlogging classification early warning algorithm. In order to verify the navigation algorithm designed in this study, we will develop a prototype system by combining Mapbox GL and AutoNavi map API.

The classification early warning algorithm of navigation route waterlogging designed in this study can provide meteorological disaster early warning information on the navigation route for the existing commercial navigation map software, and can also help the urban road traffic vulnerability evaluation research.

Acknowledgement

This research was funded by the National Key R&D Program of China (2021YFE0112300), National Natural Science Foundation of China (NSFC) (No. 41871371).

References

YIN, J., YU, D.P., WILBY, R., 2016: Modelling the impact of land subsidence on urban pluvial flooding: A case study of downtown shanghai, china. - Science of the Total Environment, 544: 744-753.

SOKOL, Z., SZTURC, J., ORELLANA-ALVEAR, J., ET AL., 2021: The role of weather radar in rainfall estimation and its application in meteorological and hydrological modelling—A review. - Remote Sensing, 13(3): 351.

LI, C.M., TANG, L.S., ZHANG, G.L., 2015: Classification standard of urban waterlogging. - China Meteorological Press, Beijing, 10.

Accuracy Enhancement of Cadastral Boundary Marker Coordinates with Smartphone Crowdsourcing

Pyry Kettunen and Mikko Rönneberg

Department of Geoinformatics and Cartography (FGI-GEOINFO) •

Finnish Geospatial Research Institute (FGI, NLS) • Vuorimiehentie 5 • 02150 Espoo

Tel.: +358 50 443 2958 • E-Mail: {firstname.lastname}@nls.fi

Keywords: smartphone, positioning, accuracy, crowdsourcing, cadastre

Summary: *Enhancing the accuracy of boundary markers in the Finnish cadastral index map was studied with reference positions by RTK GNSS and with crowdsourced positions by a smartphone game. Least accurate marker coordinates were importantly improved by smartphone positioning but additional correction methods are required to reach high-enough accuracies for practically significant national mapping by crowdsourcing. Considerable amounts of participating citizens and their measurements show high potential for the future of smartphone crowdsourcing for national mapping agencies.*

Introduction

The Finnish cadastre and its physical border marker monuments have a long history dating back to the early 21st century – for some monuments, centuries earlier. The cadastral index map is the digital representation of physical boundary monuments in the terrain. The map contains over 13 million digital border markers and is provided as open data. However, three millions of the markers have low spatial accuracy of more than one meter. The inaccuracy causes border issues for the users of the cadastral index map (Rönneberg & Kettunen, 2021).

To overcome these issues, the National Land Survey of Finland (NLS) is looking for ways to improve the quality of the cadastral index map. Promising results are coming from a gamified crowdsourcing (Gómez-Barrón et al., 2016) platform for ameliorating the cadastral index map. The platform, called Pyykkijahti (Marker Quest), is a mobile web-map based game that citizens play. In the game, citizens can either measure border marker locations with smartphone positioning or mark them missing using their mobile device. The study has two focus points: crowdsourcing (Rönneberg & Kettunen, 2021) and location accuracy enhancement (Kontiokoski, 2021), the latter of which is presented here.

Enhancement of location accuracy was studied in relation to 118 reference border markers positioned by professional surveyors and equipment of the NLS using real-time kinematic (RTK) GNSS measurements. The reference measurements were compared to smartphone measurements on the same markers, repeated on a marker up to nine times and averaged over repetitions. 26 visually clear outlier measurements were manually removed from the analysis, which decreased the positiong error by 2,37 m on average. Differential GNSS correction was tried in order to improve the positioning accuracy of smartphones but, probably due to suboptimal correction method and low precision of smartphone positioning chips, the accuracy did not improve and the correction was not considered in the analysis. Thus, the results reflect the pure overall positioning accuracy of citizens' smartphones.

322 smartphone measurements were carried out on the 118 reference markers. Their overall average error of positioning accuracy was a considerably high 4,18 m with a high standard deviation of 3,64 m. However, increasing repetitions decreased the accuracy importantly (see Fig. 2). In comparison to the coordinates of the current cadastral index map, positions of boundary markers got more accurate by 0,92 m on average. However, accuracy enhancement occurred only for markers with originally low accuracy in the cadastral index map (Tab. 1). 34 markers had a position accuracy lower than 5 m and, for these, clearly more than 90% of marker coordinates with accuracy error higher than 5 m were enhanced by meters with the smartphone positioning of the Marker Quest.

This study on positioning accuracy of citizens' smartphones revealed a relatively low overall accuracy but showed potential means for importantly higher accuracy with crowsourcing. Averaging over repeated measurements can lower the accuracy error to the meter level with enough repetitions. Introducing optimal positioning correction method and an *a posteriori* correction with satellite positions can improve the accuracy to practically significant levels in many surveying purposes. With 4 500 players and about 21 000 measurements in Jun-Oct 2021, smartphone-based crowdsourcing appears as a highly potential future method for citizen-aided complementary surveying for national mapping agencies.

Fig. 1: Crowdsourcing game Marker Quest has made a high number of citizens to measure cadastral border marker monuments.

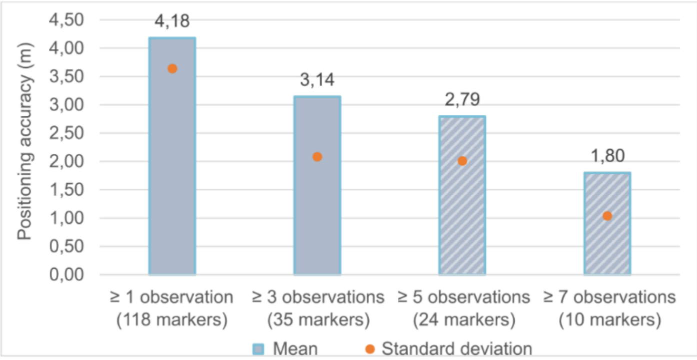


Fig. 2: Averaging over multiple measurements increases smartphone positioning accuracy. Results for ≥ 5 and ≥ 7 observations are not statistically significant because of too few observations (modified with permission from Kontiokoski, 2021).

Positioning accuracy in the cadastral index map	Border markers (pcs)	Measurement accuracy error mean (m)	Error in the cadastral index map mean (m)	Enhancement of accuracy mean (m)	Proportion of enhanced markers
< 5 meters	84	3,80	1,61	-2,19	21 %
5–10 meters	19	3,55	7,07	3,52	95 %
> 10 meters	15	7,11	22,13	15,02	93 %

Tab. 1: Least accurate border marker coordinates in the cadastral index map are enhanced importantly with smartphone positioning.

Acknowledgement

We are grateful for all the citizens who have used and continue to use Marker Quest eagerly.

References

GÓMEZ-BARRÓN, J.-P., MANSO-CALLEJO, M.-Á., ALCARRIA, R., & ITURRIOZ, T., 2016: Volunteered Geographic Information System Design: Project and Participation Guidelines. – ISPRS International Journal of Geo-Information, Volume 5 (7): 108.

KONTIOKOSKI, A., 2021: Rajamerkkien sijaintitarkkuuden parantaminen joukkoistetuilla älypuhelinmittauksilla. (Enhancing Location Accuracy of Boundary Markers by Crowdsourced Smartphone Positioning). – Bachelor's Thesis, Lapland University Of Applied Sciences.

RÖNNEBERG, M. & KETTUNEN, P., 2021: Enabling citizens to refine the location accuracy of cadastre boundary markers by gamified VGI. – Abstracts of the ICA, Volume 3: 252. https://doi.org/10.5194/ica-abs-3-252-2021

Trip and transportation mode detection using smartphone application tracking data

AGO TOMINGA, SIIRI SILM, AGE POOM AND TIIT TAMMARU

Department of Geography • University of Tartu • Vanemuise 46 • 50410 Tartu, Estonia

E-Mail: ago.tominga@ut.ee

Keywords: smartphone tracking, transportation mode detection, GPS

Abstract

Numerous methodologies have been proposed in recent years to identify transportation modes based on GPS tracking data (Anda et al., 2017). Different methodologies use various GPS and external input variables, such as velocity, acceleration, GPS signal quality or transportation and bus networks, and algorithmic approaches to detect transportation modes (Sagedhian et al., 2021). If the methodologies require specific input data, e.g., labelled data from the region from which data were collected, their replication to other data sets and spatial contexts may be challenged (Prelipcean et al., 2017). This may be especially troublesome if GPS tracking device also collects data during the time the device is not moving because in that case the successfulness of transportation mode detection also depends on successful stop and trip segmentation.

Our main research goal is to detect daily mobility profiles of transportation use. We provide a stepwise methodology for a combined trip- and transportation mode detection using rule-based and machine learning based labelling methods on unlabelled GPS data. Figure 1 presents the general workflow of our methodological workflow

First, we detect trips, which are defined as the whole journey between two stops, and split each trip into components, i.e. triplegs and segments (Figure 1A). Segments are short paths within the trip. Each time there is a short stop within the trip surpassing a certain temporal threshold (20 seconds), a new segment is formed. Following segments with similar velocity characteristics (mean speed and 95th percentile speed) within the trip are merged into triplegs. Transportation mode labelling takes place on triplegs, which enables us to include several transportation modes within the same trip. E.g., a multimodal trip may consist of a walking tripleg from origin to a bus stop, a public transport tripleg in between, and an electric scooter/bike tripleg from a bus stop to the final destination.

Second, we use a rule-based classification schema for the initial transportation mode detection (Figure 1B). We start with identitying either motorized (public transport, car) or active (walk, bike, e-scooter) modes of transportation based on velocity characteristics. We continue with comparing motorized-labelled data with general transit feed specification (GTFS) data to label public transportation use. Lastly, we use logical rules to restrict certain transitions from one transportation type to another within the same trip (e.g., it is unlikely that a person would shift several times from car to public transportation and vice versa during the same trip) and reclassify some triplegs, if needed.

Third, we use machine learning to label those triplegs that were not distinguishable with rule-based classification alone based on their similarity to already labelled data. We implement random forest algorithm (Kuhn, 2008) using 10^{th} and 90^{th} percentile, median and standard deviation of velocity and information on following/preceeding transportation type of triplegs.

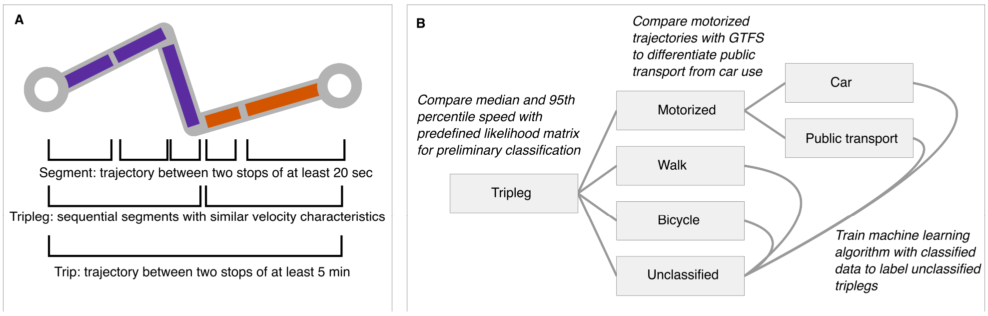


Fig. 1: The rules for trip segmentation (A) and transportation mode detection (B).

We implement our methodology on GPS-equipped mobility data, which have been collected via a research-oriented smartphone application MobilityLog (Linnap and Rice, 2014; Poom, 2019, IMO 2022). Data gathered with the application has already been used to understand mobility behaviour changes during the Covid-19 pandemic (Järv et al., 2021). Our data sample covers mobility data from more than 100 people living in Tallinn, the capital of Estonia, over a 13-month-period from October 2020 to November 2021. The mobility data set includes more than 100 million GPS points and is enriched with additional survey data about the socio-demographic characteristics, meaningful locations and lifestyle choices of study respondents. This enables us to study the longitudinal travel patterns, the regularity and stability of travel mode choice, and its eventual dynamics over time.

We use sequence analysis to detect travel chains and clustering to detect daily mobility profiles. The results including mobility profiles are later linked with survey answers regarding characteristics, such as people's gender, education level or mother tongue.

Acknowledgement

The work was supported by Estonian Research Council [PUT PRG306 and MOBTP1003]; European Regional Development Fund via the Estonian Science Infrastructure Road Map [Infotechnological Mobility Observatory (IMO)] and the Archimedes Foundation via the ASTRA program [2014-2020.4.01.16-0027].

References

ANDA, C., ERATH, A. & FOURIE, P., 2017: Transportation modelling in the age of big data. – International Journal of Urban Sciences, Volume 21 (supp. 1): 19–42.

IMO 2022. Smartphone application MobilityLog data. Infotechnological Mobility Observatory: https://imo.ut.ee/en/infrastructure/nutitelefoni-rakenduse-mobilitylog-andmed/

JÄRV, O, TOMINGA, A, MÜÜRISEPP. K. & SILM, S. 2021 The impact of COVID-19 on daily lives of transnational people based on smartphone data: Estonians in Finland – Journal of Location Based Services, Volume 15(3): 169–97.

LINNAP, M., & RICE, A., 2014: Managed Participatory Sensing with YouSense. – Journal of Urban Technology, Volume 21(2).

POOM, A, 2019: Tegevusruumiuuringud GPSi andmete abil: mobiilsusuuringute labori andmeinfrastruktuur. – In T. Pae & Ü. Mander (Eds.), Publicationes Instituti Geographici Universitatis Tartuensis: Uurimusi eestikeelse geograafia 100. Aastapäeval. Volume 113: 2014–214. Tartu Ülikooli Kirjastus, Tartu, Estonia.

PRELIPCEAN, A., GIDÓFALVI, G. & SUSILO, O., 2017: Transportation mode detection – an in-depth review of applicability and reliability. – Transport Reviews, Volume 37 (4): 442–464.

SADEGHIAN, P., HÅKANSSON, J. & ZHAO, X., 2021: Review and evaluation of methods in transportation mode detection based on GPS tracking data. – Journal of Traffic and Transportation Engineering, Volume 8 (4): 467–482

Lifting geographic relevance to the next generation location-based services in a digitally transformed world

TUMASCH REICHENBACHER AND DONATELLA ZINGARO

Department of Geography • University of Zurich • Winterthurerstr. 190 • 8057 Zurich

E-Mail: tumasch.reichenbacher@geo.uzh.ch, donatella.zingaro@geo.uzh.ch

Keywords: Geographic Relevance, digital transformation, LBS

Summary: *We present an extended conceptual framework for assessing geographic relevance in the context of digital transformation.*

Introduction

With this abstract, we want to outline the opportunities and challenges of an evolved conceptual framework for assessing geographic relevance in the context of digital transformation and contribute to a debate on future ways to enhance the experience of LBS. Even though mobile devices are pervasive in modern life, map apps and LBS which are the primary source of geographic information for navigating, spatial decision-making, and problem-solving, have not advanced substantially since their origin at the beginning of the millennium (Reichenbacher, 2019) and the underlying mapping paradigm has stayed unchanged.

A lack of adaptation to individuals and contexts can lead to problems in sense-making and usability for mobile users. Geographic relevance (GR) has been proposed as a concept to improve the utility and usability of LBS (e.g., De Sabbata and Reichenbacher, 2012; Raper, 2007). Reichenbacher et al. (2016) developed and evaluated a model of GR that demonstrated how POIs could be dynamically selected and adapted for mobile activities dependent on space-time constraints.

However, considering the fast acceleration pace of technologies, this ten-year-old model needs a revisit and extension for mainly two reasons:

1. Digital transformation is a disruptive process that aims to make physical things accessible by digital technology to link the virtual and physical world. New IT, including big data, advanced analytics, cloud computing, AI, sensors, and 5G as drivers of digital transformation (Dangi et al., 2022; Tao et al., 2019), offers great potential in information sourcing, computing, and assessing GR.
2. We now have access to sensors, digital infrastructures, and real-time information to capture user behaviour as an expression of interacting with these technologies. This offers new opportunities for supporting mobile user activities on even finer-grained levels. Thus, we see a clear need for new solutions to design adequate GR displays to support mobile citizens' digital lives.

Digital transformation allows for bridging and interfacing the physical and digital world and including non-tangible information from the digital world (Hudson-Smith and Batty, 2022). With these fundamental changes, the early approach to GR falls short and needs an extension to address the peculiarities of a digitally transformed world. The suitability of map interfaces for helping mobile users understand GR informed by digital infrastructures within an urban context, especially in the light of digital transformation processes and infrastructures, has not been fully demonstrated within the LBS community. We identify three fields for extending the GR that could profit most from digital transformation and where we see an urgent need for further research:

From static to dynamic geographic relevance

Since context information can increasingly use sensor data, context modelling and then geographic relevance modelling can also include real-time data feeds (Fabrikant, 2022). Sensing may occur locally (e.g., with Smartphone sensors) or in a sensor network (e.g., IoT). Of particular importance is real-time user behaviour in the physical, i.e., geographic space (mobility), and in the digital space (interaction with apps, databases, websites etc., through Smartphones as mediators and interfaces).

Context input from the physical and the digital world

In addition to context information sensed from the physical environment, GR modelling can also include a digitally mediated context (e.g., information about places from social networks; mobile communication at places or about places, such as text messages, social network posts, or tweets). Moreover, GR can use digital context, i.e., user behaviour in the digital world. Location can act as a connector between the digital and the physical. Physical entities and their representations in the digital world can mutually serve as contextual information (Bartling et al., 2022). They can also be input into an extended model of GR.

Geographic relevance in geographic space, digital space, and hybrid space

We still know little about physical and digital context correspondence, i.e., which parts are crucial for mutual representation. Moreover, since many activities that users have performed and could only perform in physical space are now increasingly shifted to the digital world, the geographic relevance has to be assessed from physical context factors and needs to include the digital context. This can be user behaviour sourced from the smartphone's interactions as an interface to the digital world. Also, the interplay of activities in the hybrid space (the overlap of physical and digital space) is a new field for GR (Zingaro and Reichenbacher, 2022).

Eventually, an extended GR model can use and integrate such data of users' digital behaviour on their smartphones in addition to the mobility factors and the geographic environment. This will help to improve the GR model and finally allow for better visualising geographic relevance in LBS and mobile map apps based on real-time user behaviour in physical and digital space.

References

Bartling, M., Resch, B., Reichenbacher, T., Havas, C.R., Robinson, A.C., Fabrikant, S.I., Blaschke, T., 2022. Adapting mobile map application designs to map use context: a review and call for action on potential future research themes. *Cartography and Geographic Information Science*. 1–15. https://doi.org/10.1080/15230406.2021.2015720

Dangi, R., Lalwani, P., Choudhary, G., You, I., Pau, G., 2022. Study and Investigation on 5G Technology: A Systematic Review. *Sensors* 22. https://doi.org/10.3390/s22010026

De Sabbata, S., Reichenbacher, T., 2012. Criteria of geographic relevance: An experimental study. *International Journal of Geographical Information Science* 26, 1495–1520. https://doi.org/10.1080/13658816.2011.639303

Fabrikant, S., 2022. Neuro-adaptive LBS: A research agenda on human- and context-adaptive mobile maps for pedestrian navigation and spatial learning, in: *Proceedings LBS 2022*. Munich, Germany.

Hudson-Smith, A., Batty, M., 2022. Ubiquitous geographic information in the emergent Metaverse. *Transactions in GIS* 26, 1147–1157. https://doi.org/10.1111/tgis.12932

Raper, J., 2007. Geographic Relevance. *Journal of Documentation* 63, 836–852.

Reichenbacher, T., 2019. Re-visiting Fundamental Principles of Mobile Cartography, Position paper Joint ICA Workshop on User Experience Design for Mobile Cartography: Setting the Agenda. Beijing, China. *https://use.icaci.org/Wp-Content/Uploads/2018/11/Reichenbacher.pdf*.

Reichenbacher, T., De Sabbata, S., Purves, R.S., Fabrikant, S.I., 2016. Assessing geographic relevance for mobile search: A computational model and its validation via crowdsourcing. *Journal of the Association for Information Science and Technology* 67. https://doi.org/10.1002/asi.23625

Tao, F., Qi, Q., Wang, L., Nee, A.Y.C., 2019. Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison. *Engineering* 5, 653–661. https://doi.org/10.1016/j.eng.2019.01.014

Zingaro, D., Reichenbacher, T., 2022. Modelling and communicating geographic relevance in a digitally transformed world using a digital twin, in: *Proceedings EuroCarto 2022* (accepted).

An effective and efficient neural network for fine-grained citywide crowd information prediction

XUCAI ZHANG AND HAOSHENG HUANG

Department of Geography • Ghent University • Krijgslaan 281 • 9000 Gent, Belgium

Tel.: +32 (0)9 264 47 86 • E-Mail: haosheng.huang@ugent.be

Keywords: Crowd Information, Convolutional Neural Network; k-Nearest Neighbor; Gated Recurrent Unit; Training Time Cost

Summary: *Modelling and forecasting citywide crowd information (e.g., crowd volume of a region, the inflow of crowds into a region, outflow of crowds from a region) at a fine spatio-temporal scale is crucial for urban and transport planning, city management, public safety, and traffic management. However, this is a challenging task due to its complex spatial and temporal dependences. Recently, more and more complex predictive models (e.g., with more hidden layers, sophisticated structures, or supplement information) have been proposed in the literature. However, they often suffer from the problem of high training time cost, and thus more computational resources are required and more energy is consumed, which is unfriendly to the limited computational resources or users who just expect a good accuracy under limited time cost. How to reduce training time cost while maintaining excellent predictive accuracy is still an open research challenge.*

To tackle this open research challenge, this study proposes a novel and efficient neural network model for forecasting crowd information in citywide environments, with the aims to reduce the training time cost while maintaining a better predictive accuracy than the baseline models. The proposed model (see Figure 1 for its architecture) combines recurrent neural networks (i.e., GRU) and convolutional neural network (CNN) to jointly capture the complex spatio-temporal dependences of crowd information. More importantly, a k-nearest neighbors (k-NN) module, which is shown to be an effective and efficient conventional predictive method in the literature, is added to the model to further capture more 'neighborhoods' features and accelerate the convergence of the loss function, thus reducing the training time cost of the proposed model and improving its predictive accuracy.

The evaluation with two different datasets in two different cities shows that compared to the state-of-the-art baselines, our model has better predictive accuracy (reducing the mean absolute errors MAEs by 20.99% on average) and a lower training time cost (reducing the time cost to only 26.16% on average of that of the baselines). Our model also has better abilities in making accurate predictions with low time cost under the influences of large-scale special events (when massive crowds of people are gathering in a short time) and for regions with high and irregular crowd changes. In summary, our model is an effective, efficient, and reliable method for forecasting citywide crowd information at a fine spatio-temporal scale, and has a high potential for many applications, such as city management, public safety, and transportation.

This presentation is based on the following publication: Xucai Zhang, Yeran Sun, Fangli Guan, Kai Chen, Frank Witlox, Haosheng Huang (2022), Forecasting the crowd: An effective and efficient neural network for citywide crowd information prediction at afine spatio-temporal scale. Transportation Research Part C: Emerging Technologies. DOI: https://doi.org/10.1016/j.trc.2022.103854.

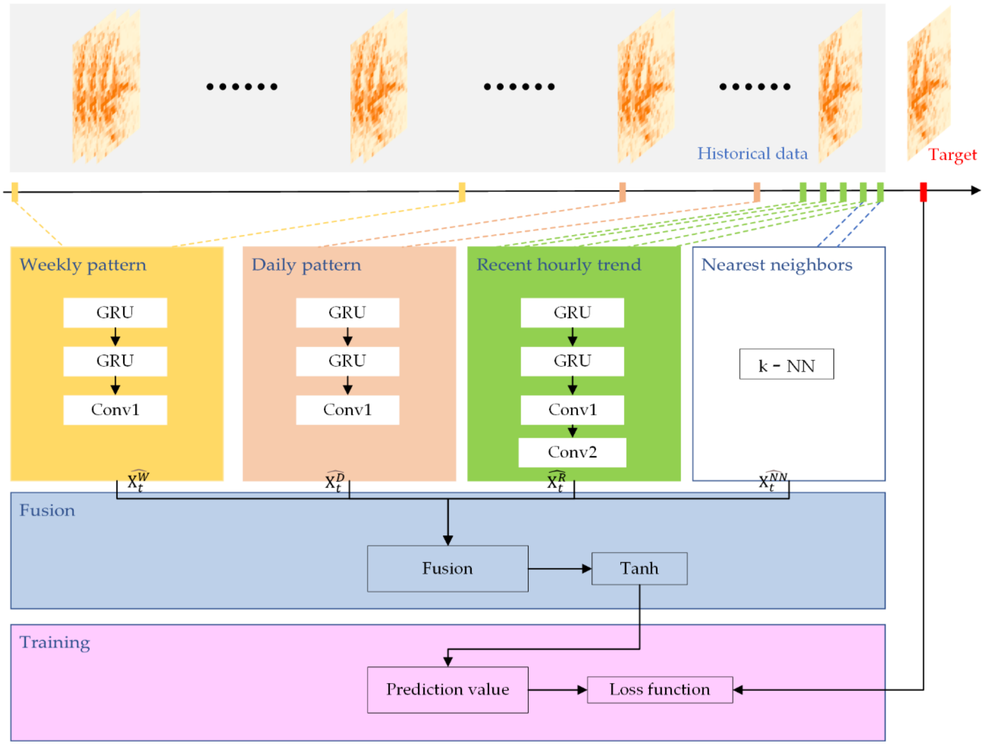


Fig. 1: Architecture of the proposed prediction method. GRU: gated recurrent unit; Conv: Convolution; k-NN: k-nearest neighbors.

A pilot service for sharing obfuscated personal level location data

VILLE MÄKINEN, ANNA BRAUER AND JUHA OKSANEN

Department of Geoinformatics and Cartography • Finnish Geospatial Research Institute, National Land Survey of Finland • Vuorimiehentie 5 • 02150 Espoo, FINLAND

Tel.: +358 40 7753 116 • E-Mail: ville.p.makinen@nls.fi

Keywords: Trajectory, personal data, privacy, open data

Summary: *There is vast and growing data of citizens' personal level movement data that could be used for example to improve the cyclability of cities and thus mitigate climate change. We present a pilot of an open movement data repository to which citizens can donate their movement data in a privacy-preserving manner.*

Introduction

The EU Commission's Climate Target Plan to cut greenhouse gas (GHG) emissions by at least 55% by 2030 leads the way for Europe to become climate neutral by 2050 (Climate Action). In recent years, more than 20% of the GHG emissions have originated from land transport (ICCT 2021). Reaching the ambitious goal of climate neutrality would need a revolution in present day thinking of personal mobility and it would also need invention and utilisation of a number of technological disruptions leading the way to more efficient use of natural resources, but also a better cost-efficiency road infrastructure investments. One of such innovations would be the open use of personal level movement data without compromising privacy.

Smart devices equipped with Global Navigation Satellite System (GNSS) receivers are ubiquitous in modern society. There are numerous applications that citizens are using to record their movement. Such trajectory data would provide invaluable insights into the non-motorized traffic in a city. Heatmaps of a number of service providers (e.g. Strava, Endomondo, Suunto) have already shown their value for traffic planning and the method provides a pragmatic solution to utilise spatially aggregated trajectory data without compromising privacy (Oksanen et al. 2015). More versatile services, such as Strava Metro (Strava 2022), are available, but access is limited only to *"organizations that plan, own, or maintain active transportation infrastructure or seek to positively influence planning processes"* leaving e.g. start-ups and researchers out, who could provide innovative uses of the data.

Analysing detailed trajectory data provides supreme application potential over commonly used heatmaps. For example, it would be possible to determine the most popular routes between regions and on which sections of the routes cyclists can travel with their desired speed and on which their travel is slowed down. One could even study what kind of turns cyclists make on specific street junctions and identify the junctions with high potential for collisions. These kinds of analyses would provide completely new input for improving the cyclability in cities.

The big ethical problem with such trajectory data is that it is highly personal and sharing citizens' movement data would compromise their privacy. For example, if a citizen has recorded their commuting trip, pinpointing their home and work place locations is often trivial. Therefore, it is important to adhere to the General Data Protection Regulation (GDPR) of the EU.

Currently the gathered data is spread among different actors. Many are commercial and do not provide any practical method to access the data, even if the data subjects have set their data public. The end result is that citizens record their movement and the data grows constantly, but there is no practical way to utilize it and the innovation potential of the data remains locked.

This situation motivated us to develop our own personal trajectory sharing service. What differentiates our approach from the others is the data ownership. Our aim is to obfuscate the donated trajectories and share them as open data. The challenge is to conserve the utility of the data but at the same time respect the privacy of the participants.

Technical aspects of the pilot service

The service (Figure 1), built using open source libraries and programs, consists of four modules: 1) Donation module, 2) Obfuscation module, 3) Statistics module, and 4) Open sharing module. In the Donation module a user can donate trajectories by uploading gpx files on the donation page. The Obfuscation module removes personal identifiers (incl. spatial and temporal aspects) from the data. The Statistics module provides users descriptive statistics of their uploaded data. Finally, the Open sharing module wraps the obfuscated trajectories into a downloadable format for registered users. For registration, the users are required to create an account and provide a working email address to use the service and explicitly give their consent before they can donate any trajectories. Email also provides a way for users to request their data to be deleted.

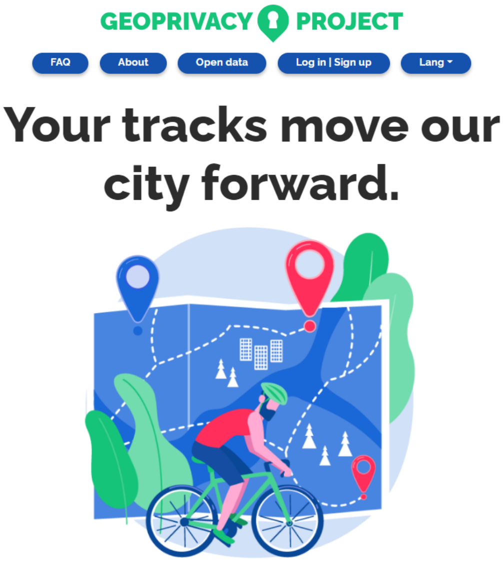


Figure 1. Entrance of the Geoprivacy pilot service for sharing personal level location data.

The donation process

When a user decides to upload her trajectories to the service, the backend will analyse and process the data and return three versions of obfuscated trajectories for each of the original trajectory. Then the user can view the results on a map and decide which versions (if any) they are comfortable with donating to the open repository.

Obfuscation methods

When the backend receives data to be processed, we will use two methods for obfuscation.

First, we truncate the trajectory based on the buildings near the endpoints of the trajectory and on the possible long stops along the trajectory using the S-TT algorithm (Brauer et al. 2022). It requires comprehensive building data from the area where we are performing the truncation, and therefore at the pilot stage it is only possible to donate trajectories that are located in Finland.

Second, we shift the truncated trajectories in time to match the first trajectory point to the closest one of the predefined, evenly spaced times of the day. This makes it more difficult to link the obfuscated trajectories to some external data, e.g. surveillance camera footage, but still allows to study general temporal aspects of the trajectory.

Next steps

We are building a service, where citizens can obfuscate their personal movement trajectories and donate them into an open trajectory database. The service is scheduled to launch this Autumn 2022. At first, only data within the borders of Finland will be considered.

Acknowledgement

The Geoprivacy project is funded by the Finnish Cultural Foundation. In addition, we made use of geocomputing platform provided by the Open Geospatial Information Infrastructure for Research (Geoportti, urn:nbn:fi:research-infras-2016072513) funded by the Academy of Finland, CSC – IT Center for Science, and other Geoportti consortium members.

References

BRAUER, A., FORSCH, A., MÄKINEN, V., OKSANEN, J. & HAUNERT, J.-H., 2022: My home is my secret: concealing sensitive locations by context-aware trajectory truncation – International Journal of Geographical Information Science, in press, doi:10.1080/13658816.2022.2081694.

CLIMATE ACTION: European Commission – 2050 long-term strategy. https://ec.europa.eu/clima/eu-action/climate-strategies-targets/2050-long-term-strategy_en (accessed 05.08.2022).

ICCT, 2021: The International Council on Clean Transportation – TRANSPORT COULD BURN UP THE EU'S ENTIRE CARBON BUDGET. https://theicct.org/transport-could-burn-up-the-eus-entire-carbon-budget/ (accessed 05.08.2022).

OKSANEN, J., BERGMAN, C., SAINIO, J. & WESTERHOLM, J., 2015: Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data – Journal of Transport Geography, Volume 48: 135-144.

STRAVA, 2022: Better cities for cyclists and pedestrians. https://metro.strava.com/ (accessed 05.08.2022).

Self-Localization Accuracy of Instrumented Probe Bicycles

HANNAH WIES AND MORITZ BEEKING

Mobility and Transport Analytics • Salzburg Research • Jakob Haringer Straße 5/3 • 5020 Salzburg

Tel.: +43 (0)662 2288-314 • E-Mail: hannah.wies@salzburgresearch.at

Tel.: +43 (0)662 2288-312 • E-Mail: moritz.beeking@salzburgresearch.at

Keywords: self-localization, instrumented probe bicycle, localization accuracy

Cycling as a healthy, sustainable mean of transport has been getting increasing political attention throughout the last decade, for example in the European green deal (European Commission 2021). The goal of increasing the share of cycling among different transportation modes requires extensive data on cycling infrastructure (Lee et al. 2014). Furthermore, as the real and perceived dangers of riding a bicycle in traffic are a major deterrent against cycling (Ng, Debnath, and Heesch 2017), the development of car-like assistance and especially warning systems is a logical consequence. Development of such systems requires data on bicycle positions and optimally their surroundings. So-called instrumented probe bicycles fitted with a variety of sensors present an effective tool to collect the needed data for both use cases. One of the most frequently pursued use case analyses accelerometer or inertial measurement unit (IMU) data for detecting infrastructure surface quality (Bíl, Andrášik, and Kubeček 2015; Neto et al. 2018). In recent years, the built-in IMUs of smartphones are used more than dedicated sensor kits (Wijerathne et al. 2018; Zang et al. 2018). Systems aiming at a comprehensive analysis of the infrastructure often incorporate cameras (Yamanaka, Xiaodong, and Sanada 2013; Nuñez, Bisconsini, and Rodrigues da Silva 2020). Some attempts at developing warning systems utilize laser-based range finders (Jeon and Rajamani 2019) or radar (Englund et al. 2019) for detecting vehicles. More often, LiDAR sensors are used to this end (Dozza et al. 2016; Van Brummelen et al. 2016; Xie, Jeon, and Rajamani 2021). To profit from recent advances in vehicle-to-everything (V2X) cooperative intelligent transport systems (C-ITS) communication, probe bicycles need to support the corresponding messages (Casademont et al. 2019).

One sensor shared by all of these systems are localization sensors (e.g. global navigation satellite system (GNSS) receivers). For infrastructure-segment-wise data collection, road accuracy with a localization error below five meters, is sufficient as showcased by multiple systems from literature (Wijerathne et al. 2018; Zang et al. 2018; Kranzinger and Leitinger 2021). For assistive and warning systems, a required localization accuracy of less than one meter is reported (Dardari et al. 2017; Miah et al. 2020). However, we consider the ability to match probe bicycles to cycle lanes with widths around one meter as necessary and thus define cycle lane accuracy as a localization error below 0.5 meters. To the best of our knowledge, a localization of single infrastructure elements like manhole covers or potholes does not exist in literature. To this end pothole accuracy with a localization error below 0.1 meters is desirable.

Fig. 1: fLTR Smartphone, XSens, Holoscene X (photos: own, Boréal Bikes)

To assess the suitability of different systems for different use cases, we compared the self-localization accuracy of three GNSS receivers depicted in fig. 1: a Xiaomi Mi9, a smartphone with an IMU-supported dual frequency GNSS; an XSens MTi 680G RTK, an RTK-corrected and INS-supported multi-frequency GNSS; and a u-blox ZED-F9P GNSS receiver. The latter receiver is integrated into the Holoscene X by Boréal Bikes, an instrumented probe bicycle fitted with multiple LiDARs, cameras, IMUs and a C-ITS communications on-board unit. The smartphone is mounted on the handlebar and the XSens on the helmet of the cyclist. The experiments are conducted in an urban as well as in a rural environment and comprise test drives in surroundings with high buildings, an underpass and without buildings. Each of these three scenarios is repeated six times at three different points in time to account for varying satellite constellations (Štern & Kos 2018). The accuracy is assessed by measuring the centreline, longitudinal, and boundary overlapping error distance using a high-definition (HD) map as ground-truth, as suggested for assessing the self-localization accuracy of autonomous vehicles (Rehrl and Gröchenig 2021). The cyclist is asked to ride along the centreline of the bicycle lane and to stop for 5 seconds at predefined positions. The centreline error distance then refers to the deviation of the measured trajectory to the centreline of a lane-level HD map. The distance between the measured stop location and the predefined stop point represents the longitudinal error distance. The boundary overlapping error distance evaluates the lane-keeping and is deduced by the overlapping distance of the 2D bounding box of the bicycle with the bicycle lane boundaries.

We expect GNSS accuracy of smartphones to meet road accuracy, while for cycle lane accuracy an RTK-corrected and INS-supported GNSS receiver is assumed to be necessary. Whether any of the tested systems is capable of pothole accuracy needs to be evaluated. The establishment of required accuracy levels of localization systems and a corresponding evaluation method can be used to select suitable hardware for different use cases of probe bicycles. The exemplary analysis of three localization systems serves as both a test of the evaluation method and guidance on the accuracy level to be expected from different types of localization systems.

Acknowledgement

The presented work is part of the Bike2CAV project, which is funded by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (Project No. 879632). The authors would like to thank Simon Gröchenig for his prior work on this project.

References

BÍL, M., ANDRÁŠIK, R., & KUBEČEK, J., 2015: How comfortable are your cycling tracks? A new method for objective bicycle vibration measurement. – Transportation Research Part C: Emerging Technologies, 56, 415–425.

CASADEMONT, J., CALVERAS, A., QUINONES, D., NAVARRO, M., ARRIBAS, J., & CATALAN-CID, M., 2019: Cooperative-Intelligent Transport Systems for Vulnerable Road Users Safety. – 7th International Conference on Future Internet of Things and Cloud (FiCloud), 2019, Istanbul, Turkey, 141–146.

DARDARI, D., DECARLI, N., GUERRA, A., AL-RIMAWI, A., PUCHADES, V. M., PRATI, G., DE ANGELIS, M., FRABONI, F., & PIETRANTONI, L., 2017: High-Accuracy Tracking Using Ultrawideband Signals for Enhanced Safety of Cyclists. – Mobile Information Systems, 2017, 1–13.

DOZZA, M., SCHINDLER, R., BIANCHI-PICCININI, G., & KARLSSON, J., 2016: How do drivers overtake cyclists? – Accident Analysis & Prevention, 88, 29–36.

ENGLUND, C., CLASEN, H., BUI, T., LINDSTRÖM, D., & ANDERSSON, J., 2019: Radar system for bicycle - a new measure for safety. – 26th ITS World Congress, 2019, Singapore.

EUROPEAN COMMISSION, 2021: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. – https://transport.ec.europa.eu/system/files/2021-12/com_2021_811_the-new-eu-urban-mobility.pdf.

JEON, W., & RAJAMANI, R., 2019: Active Sensing on a Bicycle for Simultaneous Search and Tracking of Multiple Rear Vehicles. – IEEE Transactions on Vehicular Technology, 68 (6), 5295–5308.

KRANZINGER, S., & LEITINGER, S., 2021: Eine Bestimmung der Oberflächenqualität von Fahrradinfrastruktur durch Smartphone-Beschleunigungsdaten mithilfe des k-means++-Algorithmus. – AGIT – Journal für Angewandte Geoinformatik, 7-2021, 104–116.

LEE, A., DIAS, L., MOHANTY, S., CARVALHO, T., COMMANDEUR, J., & LOVEGROVE, G., 2014: Using instrumented probe bicycles to develop bicycle safety and comfort prediction models. – 3rd International Cycling Safety Conference, 2014, Gothenburg, Sweden.

MIAH, S., MILONIDIS, E., KAPARIAS, I., & KARCANIAS, N., 2020: An Innovative Multi-Sensor Fusion Algorithm to Enhance Positioning Accuracy of an Instrumented Bicycle. – IEEE Transactions on Intelligent Transportation Systems, 21(3), 1145–1153.

NETO, G. V. A., VIANA, J. D. F., BRAGA, R. B., & OLIVEIRA, C. T., 2018: Surfaces categorization based on data collected by bike sensors. – Proceedings of the Euro American Conference on Telematics and Information Systems, 2018, Fortaleza, Brasil, 1–7.

NG, A., DEBNATH, A. K., & HEESCH, K. C., 2017: Cyclist' safety perceptions of cycling infrastructure at un-signalised intersections: Cross-sectional survey of Queensland cyclists. – Journal of Transport & Health, 6, 13–22.

NUÑEZ, J. Y. M., BISCONSINI, D. R., & RODRIGUES DA SILVA, A. N., 2020: Combining environmental quality assessment of bicycle infrastructures with vertical acceleration measurements. – Transportation Research Part A: Policy and Practice, 137, 447–458.

REHRL, K., & GRÖCHENIG, S., 2021: Evaluating Localization Accuracy of Automated Driving Systems. – Sensors, 21(17), 5855, 1–22.

ŠTERN, A.; KOS, A., 2018: Positioning Performance Assessment of Geodetic, Automotive, and Smartphone GNSS Receivers in Standardized Road Scenarios. – IEEE Access, 6, 41410–41428.

VAN BRUMMELEN, J., EMRAN, B., YESILCIMEN, K., & NAJJARAN, H., 2016: Reliable and low-cost cyclist collision warning system for safer commute on urban roads. – IEEE International Conference on Systems, Man, and Cybernetics, 2016, Budapest, Hungary, 003731- 003735.

WIJERATHNE, N., KV, S., MARAKKALAGE, H., BELTRAN, V., YUEN, C., & LIM, H., 2018: Towards comfortable cycling: A practical approach to monitor the conditions in cycling paths. – IEEE 4th World Forum on Internet of Things, 2018, Singapore, 783–788.

XIE, Z., JEON, W., & RAJAMANI, R., 2021: Low-Density Lidar Based Estimation System for Bicycle Protection. – IEEE Transactions on Intelligent Vehicles, 6(1), 67–77.

YAMANAKA, H., XIAODONG, P., & SANADA, J., 2013: Evaluation Models for Cyclists' Perception Using Probe Bicycle System. – Proceedings of the Eastern Asia Society for Transportation Studies, 10, 1413–1425.

ZANG, K., SHEN, J., HUANG, H., WAN, M., & SHI, J., 2018: Assessing and Mapping of Road Surface Roughness based on GPS and Accelerometer Sensors on Bicycle-Mounted Smartphones. – Sensors, 18, 914, 1–17.