ARTICLE IN PRESS

Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Weather effects on human mobility: a study using multi-channel sequence analysis

Vanessa S. Brum-Bastos*, Jed A. Long, Urška Demšar

School of Geography & Sustainable Development, University of St Andrews, Irvine Building, North Street, St Andrews, UK

ARTICLE INFO

Keywords:
Context-aware movement analysis
Context-aware similarity
Human mobility
Human movement
Multi-channel sequence analysis
Context

ABSTRACT

Widespread availability of geospatial data on movement and context presents opportunities for applying new methods to investigate the interactions between humans and weather conditions. Understanding the influence of weather on human behaviour is of interest for diverse applications, such as urban planning and traffic engineering. The effect of weather on movement behaviour can be explored through Context-Aware Movement Analysis (CAMA), which integrates movement geometry with its context. More specifically, we use multi-channel sequence analysis (MCSA) to represent a person's movement as a multi-dimensional sequence of states, describing either the type of movement or the state of the environment throughout time. Similar movement patterns can then be identified by comparing and aligning mobility sequences. In this paper we apply CAMA and MCSA to explore weather effects on human movement patterns. Data from a GPS tracking study in a Scottish town of Dunfermline are linked to weather data and converted into multi-channel sequences which are clustered into groups of similar behaviours under specific weather typologies. Our findings show that the CAMA + MCSA method can successfully identify the response of commuters to variations in environmental conditions. We also discuss our findings on how travel modes and time spent at different places are affected by meteorological conditions, mainly wind, but also rainfall, daylight duration, temperature, comfort and relative humidity.

1. Introduction

The spread of geolocated smartphones and the decreasing price of GPS devices have contributed towards the production of large amounts of data on human movement of unprecedented spatio-temporal quality (Meekan et al., 2017). New human mobility studies attempt to link such movement data with contextual information (such as points of interest) to gather insights into, for example, commuting behaviour (Beecham, Wood, & Bowerman, 2014; Gong, Chen, Bialostozky, & Lawson, 2012), tourist behaviour (Meijles, de Bakker, Groote, & Barske, 2014; Versichele, Neutens, Delafontaine, & Van de Weghe, 2012), or retail choice decisions and human activities (Siła-Nowicka et al., 2016). However, integrating high resolution GPS trajectories and dynamic spatio-temporal contextual information remains an underexplored approach for studying the effects of weather on human movement, despite its relevance for urban planning (Givoni, 1974; Ng, 2012), traffic engineering (Dunne & Ghosh, 2013), retail planning (Thakuriah, Sila-Nowicka, & Paule, 2016), tourism (de Freitas, 2003), health (Tucker & Gilliland, 2007), psychology (Nerlich & Jaspal, 2014) and epidemiology (Horowitz, 2002).

Specific weather conditions often trigger changes in human behaviour, for example, higher temperatures increase aggressiveness (Anderson, 2001; Carlsmith & Anderson, 1979) and lower temperatures contribute to irritability and combativeness (Schneider, Lesko, & Garret, 1980; Worfolk, 1997). Different components of weather have different magnitudes of importance, for example, air temperature, direct solar radiation and wind speed have a more significant influence on human behaviour than humidity (de Montigny, Ling, & Zacharias, 2012). However, it is challenging to understand how weather influences human behaviour because the responses are partially a result of individual preferences (de Freitas, 2015). Some individuals are more responsive to the thermal component of weather, i.e. the combined effects of air temperature, humidity and solar radiation, while some are more receptive to physical components like rain, and others are more greatly affected by the aesthetic components, such as cloud coverage and sunshine. Yet, most individuals do respond to the combination of all three of these components (de Freitas, 1990).

Traditionally, these interactions have been explored through questionnaires and multidimensional scaling methods within the field of human biometeorology (Cabanac, 1971; de Freitas, 1990; Manu, Shukla, Rawal, Thomas, & de Dear, 2016). With the increased availability of tracking and environmental data we however propose that the

* Corresponding author.
E-mail addresses: vdsbb@st-andrews.ac.uk (V.S. Brum-Bastos), jed.long@st-andrews.ac.uk (J.A. Long), urska.demsar@st-andrews.ac.uk (U. Demšar).

https://doi.org/10.1016/j.compenvurbsys.2018.05.004
Received 31 January 2018; Received in revised form 15 May 2018; Accepted 15 May 2018
0198-9715/

Please cite this article as: Brum-Bastos, V.S., Computers, Environment and Urban Systems (2018), https://doi.org/10.1016/j.compenvurbsys.2018.05.004

effect of weather on movement behaviour can be explored through Context-Aware Movement Analysis (CAMA), which integrates movement geometry with its context, i.e. with the surrounding biological and environmental conditions that might be affecting movement (Andrienko, Andrienko, & Heurich, 2011; Demsar et al., 2015; Dodge et al., 2013). More specifically we use multi-channel sequence analysis (MCSA) to represent a person's movement as a sequence of states, describing either the type of movement or the state of the environment throughout time. Similar movement patterns can then be identified (termed context aware similarity analysis) by comparing and aligning mobility sequences.

Similarity analysis is one of the most common tasks in movement analytics and consists of using distance measures and grouping methods to split trajectories (Demšar et al. 2015) into groups of elements more similar amongst them than to other groups (Jain et al., 1999), which followed by clustering allows the identification of spatio-temporal movement patterns that might be linked to behaviour (Dodge, Weibel, Ahearn, Buchin, & Miller, 2016). Similarity is often established based on geometry or physical attributes; geometrical similarity solely relies on measures of spatial and temporal distances, and physical similarity relies on movement attributes such as speed, turning angle, acceleration and direction (Demsar et al., 2015). Context-aware similarity is based on multiple attributes (Andrienko et al., 2011; Buchin, Dodge, & Speckmann, 2014; Demsar et al., 2015; Sharif & Alesheikh, 2017b) describing the conditions within which the movement took place.

Context-awareness is a recent trend (Sharif & Alesheikh, 2017a), as a result there are few context-aware methods for assessing similarity between trajectories. Sharif & Alesheikh (2017b) generalized the dynamic time warping (DTW) to develop a context-based dynamic time warping (CDTW) method, which matches trajectories with contextual similarity even if they are not concurrent. This method is highly dependent on arbitrary weights for the contextual variables, restricted to numeric context and disregards changes of context between two points in time. i.e., same contexts are considered similar even when they are not concurrent. De Groeve et al. (2016) uses single channel sequence alignments and Hamming Distance to understand the temporal variation of habitat use by roe deer; the similarity is measured by the cost to transform a sequence of habitat use into another. This method is able to handle only one contextual variable at time, therefore it is not able to handle the interactive effect of multiple contextual variables on movement. Buchin et al. (2014) modified existing similarity measures to make them context-aware, more specifically they defined the distance between two points as the sum of their contextual and spatial distances. The transition costs between contexts are defined by the user and the method is restricted to contextual data in the form of polygonal divisions.

In this paper we propose to use multi-channel sequence analysis (MCSA) to perform context-aware similarity analysis (CASA) and cluster trajectories into groups of similar behaviour. MCSA is a new analysis tool for movement data where contextual information can now be readily combined with detailed tracking datasets. The main advantage of this approach is that it also is possible to consider as many channels (contextual variables) as desired at once. It is common in movement research to simultaneously consider multiple environmental variables, which makes MCSA particularly relevant for studying human mobility, traffic, transportation and wildlife ecology; areas in which movement behaviour may be contextualised by other dynamic environmental variables such as air temperature, vegetation indices, humidity, wind speed, air pollution and snow coverage. Single channel analysis has been used before to explore spatio-temporal patterns on the activity of visitors in Akko's Old city – Israel (Shoval & Isaacson, 2007) and to analyse sequential habitat use by roe deer in North-East Italy (De Groeve et al., 2016). Shoval & Isaacson (2007) focused on sequences of locations, i.e. the movement itself, while De Groeve et al. (2016) emphasized sequences of habitat use classes, i.e. the context surrounding movement. Horanont et al. (2013) looked at GPS traces from mobile phone users, coarse scale movement data, hourly temperature, rainfall and wind speed to explore the independent effects of each variable on people's activity patterns. We innovate by applying MCSA, for the very first time, to perform CAMA of fine scale human movement data to simultaneously consider movement and context by looking at the combined and single effects of six meteorological variables.

Despite the novelty of MCSA in movement research, sequence analysis has been consistently used in medical and social sciences, particularly within bioinformatics and life courses research (Idury & Waterman, 1995; Abbott 1995; Abbott & Tsay 2000). In bioinformatics, a sequence represents the DNA molecule as a string of characters (which stand for specific nucleotides), between a precise start and end point; the comparison of similarities and differences between those strings allows the identification of nucleotide sequences related to genetic diseases and traits. We propose that the same principle can be applied to movement trajectories for identifying groups of people with similar movement patterns, i.e., clusters of similar behaviour (Billari, 2001). Further, we propose to not only represent the trajectories with one sequence only, but to use Multi-channel sequence analysis (MCSA), which allows for comparison of sequences consisting of several dimensions (channels) (Gauthier et al., 2010). For this, we link data from a GPS tracking study to weather data and convert the information into multi-channel sequences in a first fully data-driven attempt to explore weather effects on human movement patterns.

The rest of the paper is structured as follows: first we describe the GPS tracking data and weather datasets used in our analysis. Next, we explain how the meteorological data sources were combined and integrated with the GPS tracking data and finally converted into sequences. Next, multi-channel sequence analysis is applied to identify changes in group movement patterns related to weather. We conclude with considerations on our findings, the potential of the methodology and ideas for future research.

2. Methodology

To study the influence of weather on human mobility behaviour we used a five-step process (Fig. 1). In Step 1, we integrate trajectories with contextual data by using trajectory annotation to link GPS points to weather variables, which resulted in contextualised trajectories. In Step 2, we transform those trajectories into multi-channel sequences by creating alphabets with codes for each weather variable, travel mode and places. In Step 3, we use optimal matching distances (Abbott & Tsay, 2000) to calculate a dissimilarity matrix describing the degree of difference between each pair of multi-channel sequences in our dataset. In Step 4, we use Ward's clustering (Murtagh & Legendre, 2011) algorithm to partition the sequences into similarity based groups, which represent groups of people showing similar movement behaviour under particular weather conditions. In Step 5, we perform statistical test to validate and understand differences between groups.

Trajectory annotation and sequencing were performed using PostgreSQL 9.4 database manager, VANJU library and its dependencies under Python 2.7, for more details refer to Brum-Bastos, Long, & Demšar (2016). The MCSA, including optimal matching distances, Ward's clustering and statistical tests, was performed using TraMineR 1.8-9 and cluster 1.14.4 libraries under R 3.4.1, for more details on the equations used by these libraries please refer to Gabadinho, Ritschard, Studer, & Müller (2009); Maechler, Rousseeuw, Struyf, Hubert, & Hornik (2018) respectively.

2.1. Movement data

We analysed a human movement dataset where GPS devices were carried by volunteers from the Kingdom of Fife – UK (Fig. 2a) (Siła-Nowicka et al., 2016). The data were collected between the 28th of September 2013 and the 10th of January 2014 as part of the GEOCROWD project (Siła-Nowicka et al., 2016), in which 6000 individuals