Mobility profiling: Identifying *scouters* in the crowd

Licia Amichi[†⋆], Aline C. Viana[⋆], Mark Crovella[‡], and Antonio F. Loureiro[×]
[†] Ecole Polytechnique [⋆] INRIA [‡] Boston University [×] Federal University of Minas Gerais

ABSTRACT

The prediction of individuals' dynamics has attracted significant community attention and has implication for many fields: e.g. epidemic spreading, urban planning, recommendation systems. Current prediction models, however, are unable to capture uncertainties in the mobility behavior of individuals, and consequently, suffer from *the inability to predict visits to new places*. This is due to the fact that current models are oblivious to the exploration aspect of human behavior. This paper contributes better understanding of this aspect and presents a new strategy for identifying exploration profiles of a population. Our strategy captures spatiotemporal properties of visits – i.e. a known or new location (spatial) as well as a recurrent and intermittent visit (temporal) – and classifies individuals as scouters (i.e., extreme explorers), routineers (i.e., extreme returners), or regulars (i.e., with a medium behavior). To the best of our knowledge, this is the first work profiling spatiotemporal exploration of individuals in a simple and easy-to-implement way, with the potential to benefit services relying on mobility prediction.

ACM Reference Format:
Licia Amichi[†⋆], Aline C. Viana[⋆], Mark Crovella[‡], and Antonio F. Loureiro[×] [†] Ecole Polytechnique [⋆] INRIA [‡] Boston University [×] Federal University of Minas Gerais . 2019. Mobility profiling: Identifying *scouters* in the crowd. In *The 15th International Conference on emerging Networking EXperiments and Technologies (CoNEXT '19 Companion), December 9–12, 2019, Orlando, FL, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3360468.3366771

1 INTRODUCTION

Many prediction models have been proposed to forecast individuals trajectories. However, they all show limited bounded predictive performance [1]. Regardless of the applied methods (e.g., Markov chains, Naive Bayes, neural networks), the type of prediction (i.e., next-cell or next place) or the used data sets (e.g., GPS, CDR, surveys), accuracy of prediction never reaches the coveted 100%. The reasons for such limitations in the accuracy are manyfold: the lack of ground truth data, human beings' complex nature and behavior, as well the exploration phenomenon (i.e., visits to never seen before places) [1, 2, 6]. In this paper, we focus on the exploration problem, which has rarely been tackled in the literature but indeed, represents a real issue [1]. By construction, most prediction models attempt to forecast future locations from the set of known places, which hinders predicting new unseen places and by consequence, reduces the predictive performance. In [5], the authors reported the existence of two mobility profiles: (i) returners and (ii) explorers, and suggested that the probability of exploring new areas is correlated with the number of frequently visited places. However, this classification can be unsuitable; for instance, a person who regularly visits two different locations and usually explores many new areas is considered to be a returner, while a person who spends most of her time between eight different locations and rarely visits new ones can be viewed as an explorer. The authors in [6] corroborate the results drawn in [5] and shown the existence of two distinct groups of individuals: (i) travelers, who move around extensively, and (ii) locals, who move in a more constrained area and revisit many of their locations. Nevertheless, they do not bring any understanding of the exploration behavior of individuals. Although their approach does not classify all individuals and results in five groups of individuals, only two groups were interpreted and considered to be significant. In [1], an exploration prediction model was proposed based on random guessing of explorations. Still, this model suggests that all individuals have the same probability to explore, which contradicts what was shown in [5, 6].

Thus, when considering the exploration problem, previous studies either did not provide any consideration of the exploration factors of individuals, or divided the population based on properties that are not always consistent, or assumed that all individuals have the same propensity to explore. Our main goal in this work is to understand the exploration phenomenon and answer the following question: *What type of visits characterize the mobility of individuals?* Using newly designed metrics capturing spatiotemporal properties of human mobility – i.e., known/new and recurrent/intermittent visits – our strategy identifies three groups of individuals according to their degree of exploration: scouters, routineers, and regulars. In the future, we plan to deeply investigate the mobility behavior of individuals in each profile and to assign to each individual an *exploration factor* describing her susceptibility to explore.

2 PROPOSED METHOD

To understand human mobility dynamics and identify the circumstances inciting individuals' propensity to break their routine and explore new spots, we divide human moves into two complementary movements: explorations and returns. Indeed, at each instant, an individual has two choices: she either walks back to a place she visited in the past, or explores a new site. Hereafter, we define (i) an **exploration** as *a visit to a never seen before location*, i.e., a location that is not present in the history of a given individual and (ii) a **return** as *a visit to a previously seen locality.*

2.1 Formalization

Let M be the Finite-State Automaton (FSA) describing an individual movements, as shown in Fig. 1, with two possible states: *exploring*

CoNEXT '19 Companion, December 9–12,2019, Orlando, FL, USA

ACM ISBN 978-1-4503-7006-6/19/12…$15.00
https://doi.org/10.1145/3360468.3366771

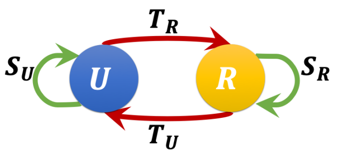

Figure 1: State Diagram of Human movements

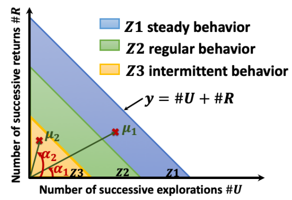

Figure 2: Successive visits

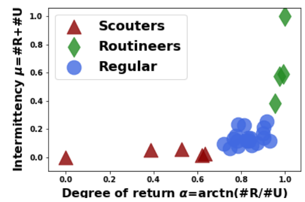

Figure 3: Privamov

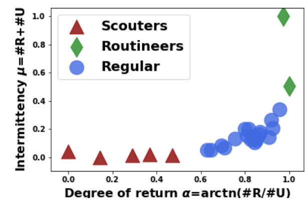

Figure 4: Macaco

(**U**) and *returning* (**R**). Two possible inputs affect such states: *return* (T_R or S_R) by going back to historically known locations, and *explore* by discovering new spots (T_U or S_U). In the **U** state, exploring new areas (S_U) has no effect and keeps the individual in the state **U**. On the other hand, moving back to a known location (T_R), though recently explored, gives M an input and shifts the state from **U** to **R**. In the **R** state visits to usual places (S_R) does not change the state, however, a discovery of a new spot (T_U), shifts the state back to the **U** state. We associate to each individual the average number of self-transitions S_U she made in the state **U** (i.e., $\#U$) and S_R in the state **R** (i.e., $\#R$).
Using the spatiotemporal footprints captured in a given dataset, we define the following metrics to describe the exploration habits of individuals:

Definition 1 (*Intermittency* μ). ***Intermittency*** *μ is the sum of the average number of movements performed in each state **U** and **R**.* ($\mu = \#R + \#U$)

Definition 2 (*Degree of return* α). ***Degree of return*** *α is the angle whose tangent is the ratio between the average number of successive visits of type **R** over the average number of successive visits of type **U*** $\left(\alpha = \text{arctg}\left(\frac{\#R}{\#U}\right)\right)$.

What do the metrics α and μ capture? The *intermittency* μ captures the transition patterns of individuals between the states **U** and **R**. The more distant an individual is from the origin, the steadier she is. When $\#U$ or $\#R$ increases the sum $\#U + \#R$ increases, indicating that fewer shifts occur between **U** and **R**. Therefore, the intermittency metric reveals whether the individual is versatile or prefers to be steady. For instance, the individual 2 (i.e., with μ_2) in zone 3 (i.e., Z3) in Fig. 2, is more intermittent than the individual 1 (i.e., with μ_1) in zone 1 (i.e., Z1). The *degree of return* reports the exploration habits of an individual compared to her returns, whether she relatively performs more explorations or returns compared to the average statistics raised from the population. In Fig. 2, individual 1 is more prone to explore than individual 2 ($\alpha_1 < \alpha_2$).

2.2 Preliminary evaluation

For each individual, we first measure her *intermittency* and *degree of return*. Next, we use the Gaussian mixture probabilistic model to investigate whether we can split the population into distinct cohesive and significant groups.

Dataset: Our first dataset source is an anonymized trace collected by the MACACO project [4] during approximately 34 months. It contains timestamped GPS-like coordinates of 99 individuals. The second dataset contains the timestamped geolocalized trajectories of 100 volunteers collected by the Privamov project [7] during 14 months. We consider only participants that appear with more than 500 measurements with at least 10 days of contiguous data and a frequency of sampling equal to 5min, resulting in 25 individuals for MACACO and 29 individuals for Privamov. In this work, we tessellate the concerned geographical regions in the datasets with grids of side 600m, which results in an assignment to each GPS coordinate, a cell with a unique identifier.

Results: Fig. 3 and 4 show that our metrics identify three distinct profiles in terms of human mobility dynamics. The first profile is *scouters or extreme explorers*, whose degree of return is relatively low and who are intermittent and constantly shifting from a state to another. These individuals are more prone to explore new areas. The second is *routineers or extreme-returners*, who have a surprisingly large degree of return and remain steady in the different states. These individuals rarely perform explorations and prefer to stick among the common and known places. Finally, *regulars* are individuals who have a medium behavior alternating between explorations and revisits. Our metrics results in a natural clustering of individuals, although having a different number of frequently visited locations, individuals who usually break their routines to explore are viewed as scouters, unlike in [5] where some can be clustered as explorers and others as returners. Contrary to [6] our approach captures three major mobility features that fully describe the exploration phenomenon: uniqueness of visits (i.e. explorations), intermittency between returns and explorations (its importance was shown in [3] as stationarity), and the ratio of explorations compared to returners and splits the populations accordingly.

2.3 Discussions and Future Work

In this study, we split the population according to their propensity to explore: *How often does an individual explore? How many new places does she visit consecutively?* This profiling resulted in three distinct classes: (i) scouters, who are more adventurous and like to discover many new places sequentially; (ii) routineers, who are more steady and rarely leave their comfort zone to explore new ones and (iii) regulars, who have a medium behavior alternating between explorations and revisits. In the future work, we aim to assess the effectiveness of our clustering method by investigating each group independently and measuring new spatiotemporal features –e.g., the duration of visits, the number of stops, the ratio of places visited only once or distances walked – and identifying the features that are specific to each mobility profile. Further, we aspire to understand the exploration phenomenon and to associate to each individual a factor that given her mobility profile and history, can tell whether she is more susceptible to return to a previously know place or perform a visit to a new region, and this can be a prime mover in improving the accuracy of prediction.

REFERENCES

[1] A. Cuttone, S. Lehmann and M. C. Gonzalez. 2018. Understanding predictability and exploration in human mobility. In *EPJ Data Science*.
[2] B. Schmidt. 2005. Human Factors in Complex Systems The Modelling of Human Behaviour. In *19th European Conference on Modelling and Simulation*.
[3] D. de C. Teixeira, A. C. Viana, M. S. Alvim, J. M. Almeida. 2019. Human Factors in Complex Systems The Modelling of Human Behaviour. In *Deciphering Predictability Limits in Human Mobility', ACM SIGSPATIAL Chicago. USA*.
[4] K. Jaffr'es-Runser, G. Jakllari, T. Peng, V. Nitu. 2017. Crowdsensing Mobile Content and Context Data: Lessons Learned in the Wild. In *Workshop on Data Analytics for Mobile Networking*.
[5] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti and A.-L. Barabási. 2015. Returners and explorers dichotomy in human mobility. In *Nature Communications volume 6, Article number: 8166*.
[6] L. Scherrer, M. Tomko, P. Ranacher and R. Weibel. 2018. Travelers or locals? Identifying meaningful sub-populations from human movement data in the absence of ground truth. In *EPJ Data Science*.
[7] S. Ben Mokhtar, A. Boutet, L. Bouzouina, P. Bonnel, O. Brette, L. Brunie, M. Cunche, S. D'Alu, V. Primault, P. Raveneau, H. Rivano, R. Stanica. 2017. PRIVA'MOV: Analysing Human Mobility Through Multi-Sensor Datasets. In *NetMob*.