Program of work during the visit

Generalized predictability measure for human mobility accounting for spatial, temporal, and sequential dimensions – GenPM-STD

Keywords: human mobility, spatiotemporal prediction, Artificial Intelligence, Computational Social Science

The availability of fine-grained location data (e.g., CDRs, GPS traces) enables the study and prediction of human movements, which are of great value in many applications such as service provision, location-based advertisements, and disease spread modeling.

The seminal paper of Song et al. [1] proposes the entropic level of individuals' movement trajectories as a theoretical upper bound on the predictability of human mobility. A regular trajectory consists mostly of repeatable routines (stationary segments) and therefore have lower entropy (more predictable). Entropy is calculated using the Lempel-Ziv algorithm [9], which scans through the trajectory and calculates the length of yet unseen patterns. The maximum predictability observed on a dataset of 45,000 mobile phones users is 93% [1].

However, Song et al.'s predictability metric, still extensively used by the human mobility community, has several limitations. First, it is strongly dependent on the length of trajectories, the number of places visited and trips between them, and the time between transitions [2-7]. Second, since it does not capture long-distance dependencies in human trajectories, actual predictability should be higher [7]. Indeed, recurrent neural networks (RNNs), which capture dependencies of varying lengths, surpass the theoretical upper bound of predictability [8]. Finally, it does not consider trajectories' space and time dimensions: two individuals with the same sequence of symbols have the same entropy, but these symbols may be associated with locations that are displaced differently in space and time, which should influence the two individuals' predictability.

The purpose of this visit will be **to develop a novel predictability metric for human mobility** that overcomes the limitations above. This new metric will consider spatial and time dimensions, long-distance dependencies, and will be dataset-independent. **The metric will be tested on public mobility data** and the consistency of the theoretic upper bound **will be tested with state-of-the-art mobility predictors**, including those based on **Artificial Intelligence (AI)**, to verify that their accuracy does not surpass the theoretical predictability upper bound. The metric will allow characterizing the performance of mobility predictors better and to design more explainable models, both questions of interest to DigiCosme. In particular, the visit's work relates to IID-3 (for its relation to ML, knowledge extraction, and reasoning) and to the ComEx-3 (for its applicability in user-aware intelligent transport and distributed networking). The newly developed methodology will be integrated into library scikit-mobility, of which the applicant is founder and an active developer.

References

[1] Song et al., Limits of predictability in human mobility, Science 327:5968, 2010.
[2] Lu et al., Approaching the limit of predictability in human mobility, Scientific Reports 3, 2013.
[3] Cuttone et al., Understanding predictability and exploration in human mobility, EPJ Data Science 7:1, 2018.
[4] Ikanovic and Mollgaard, An alternative approach to the limits of predictability in human mobility, EPJ Data Science 6:1, 2017.
[5] Smith et al., A refined limit on the predictability of human mobility, in IEEE Int. Conf. on Pervasive Comp. and Comm., 2014.
[6] Do Couto Teixeira et al., Deciphering predictability limits in human mobility, Proc. ACM SIGSPATIAL, 2019.
[7] Kulkarni et al., Examining the Limits of Predictability of Human Mobility, Entropy 21:432, 2019.
[8] Luca et al., A Survey on Deep Learning for Human Mobility, ACM Computing Surveys 55, 1, Article 7, 44, 2021.
[9] Ziv and Lempel, Compression of Individual Sequences via Variable-Rate Coding, IEEE Trans. Inf. Theory, 24:5, 1978.