Deep Generative Models for Synthesizing Mobile Traffic Data

Kai Xu
The University of Edinburgh

City-scale mobile traffic data has numerous applications within and beyond networking. Prominent examples of network problems that benefit from mobile traffic data include resource management, mobile network infrastructure planning, network energy efficiency optimization, or network monitoring. Beyond networking, such data has wide range of applications, including in urban sensing and computing, for inference of commuting patterns and segregation, monitoring demographic patterns, detection of land use and its dynamics, transportation engineering and urban planning, or road traffic surveillance.

All aforementioned applications are enabled by mobile traffic data that is inherently de-personalized and so poses reduced privacy risks from sharing. Yet, such data is deemed commercially sensitive by mobile network operators and so is not easily and widely accessible. Even the few who have access to this data have to go through restrictive Non-Disclosure Agreements (NDAs) with operators. The result is that (i) the potential of mobile traffic data to feed innovation is curbed, and (ii) current research based on mobile traffic is not reproducible or verifiable.

Motivated by the above to overcome the mobile traffic data access barrier, we aim at synthesizing dependable city-scale mobile traffic data. To this end, we take the deep generative modeling approach that has seen great success in recent years for synthetic data generation in other domains (e.g., computer vision, health, finance, art). At a high level, our approach can be viewed as leveraging access to limited amount of real traffic data to design traffic generation models that can then be relied on to generate unlimited amount of "like-real" data.

In this presentation, we will outline the suite of deep generative models we have developed till date for city scale mobile traffic synthesis, including CartaGenie for spatial traffic snapshot generation [1] and SpectraGAN for spatiotemporal traffic generation [2]. A common theme across these models is to learn mapping between publicly available static, spatial context data and traffic, which enables a flexible and generalizable mobile traffic synthesis paradigm to generate traffic for any target region given its context data.

We will not only present the key insights and design choices underlying our models for generating mobile traffic of various kinds but also highlight the challenges they address towards high-fidelity and generalizable synthesis. We will also give an overview of our extensive evaluation studies considering a wide range of quantitative and qualitative fidelity metrics as well as several downstream use cases. These evaluations are based on multi-city operator provided traffic datasets from two major European countries.

References

[1] K. Xu, R. Singh, H. Bilen, M. Fiore, M. K. Marina, and Y. Wang, "CartaGenie: Context-Driven Synthesis of City-Scale Mobile Network Traffic Snapshots," in *Proceedings of 20th IEEE International Conference on Pervasive Computing and Communications (PerCom'22)*. IEEE Computer Society, Mar 2022, pp. 119–129. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/PerCom53586.2022.9762395

[2] K. Xu, R. Singh, M. Fiore, M. K. Marina, H. Bilen, M. Usama, H. Benn, and C. Ziemlicki, "SpectraGAN: Spectrum Based Generation of City Scale Spatiotemporal Mobile Network Traffic Data," in *Proceedings of the 17th International Conference on Emerging Networking EXperiments and Technologies (CoNEXT'21)*. ACM, 2021, p. 243–258. [Online]. Available: https://doi.org/10.1145/3485983.3494844