Internship:
Harnessing the Power of Big Data for Optimizing Future Generation Transportation

Context

Accessibility [1] is a graph-theoretical metric that measures how well a certain place (aka node) is connected to the surrounding urban area (aka the rest of the transportation graph). A place enjoys high accessibility when it can be easily and rapidly reached from any other place in the area. Accessibility is a key indicator for a town or a district, as it determines quality of life, job attractiveness, value of properties. The design of public transit (bus, metro, train lines and frequencies) strongly impacts accessibility and is often the cause of spatial inequality between city centres and suburbs.

Accessibility can now be objectively quantified relying on big data, publicly available. For instance, GTFS is a data standard created by Google and now adopted by all the main transportation authorities around the world to publish their transit schedules. GTFS data can then be used by 3rd party applications to offer several services, e.g., route planner apps. Other socio-economic and demographic data are also publicly available.

The goal of this internship is, first, to create a model of current transportation in an urban conurbation (e.g. the Parisian region), based on GTFS data. Then, we will optimise such a network, I.e., modify routes and schedules, in order to reduce the gap between accessibility in the city centre and the suburbs.

Internship Activity

1st part: The student will deploy the open source big data application *CityChrone* [1] on our local computer cluster. CityChrone is able to ingest GTFS data [2], population information [3] and Open Street Maps and visualize accessibility scores.

2nd part: The student will solve a Network Design Problem [4], i.e., the construction of an optimal public transit network that is able to minimize the accessibility gap between the centre and the suburbs. Such kinds of problems are NP-hard and have been classically solved via meta-heurstics, whose limit is that they start from scratch when applied to a new network. We will instead apply Graph Embedding and Deep Reinforcement Learning [6, 7], in which we can accumulate a knowledge in the form of general network embedding parameters and Q-function approximations, which can be later reused when optimizing other networks (a feature that is called transfer learning [5]). While such AI techniques have been successfully adopted to solve canonical graph problems, we will need to adapt them to our network construction case.

Administrative information

Where: Télécom SudParis, Palaiseau campus

When: Starting from March 2022, duration: 6 months

Supervisor: Assoc. Prof. Andrea Araldo and Prof. Mounim El Yacoubi (Institut Polytechnique de Paris – Télécom SudParis)

To apply

Candidates must be M2 level (2nd year of MSc or last year of cycle ingénieur) in Data Science, Computer Science, Transportation Engineering or Applied Mathematics.

The candidate must be proficient in Python and have excellent programming skills. She/he must also be able to perform some basic server administration and work with remote servers (ssh). She/he must be proficient with github or similar platforms. A previous experience in deep learning or reinforcement learning is a big plus.

In the first interview, the student will have to show that she/he is able to create a virtual machine with Ubuntu-server (no-gui) installed there, install CityChrone on such a machine and access its graphical interface remotely or from the host. If this installation is not possible, she/he has to precisely show the blocking points.

The candidate must send to andrea.araldo@telecom-sudparis.eu: (i) marks of all her/his BSc and MSc studies, (ii) Cv, (iii) motivation letter.

Perspectives

Possibility to start a PhD, after the internship. The topic would be big data and reinforcement learning for the design of multi-modal transit.

References

[1] Biazzo, I., Monechi, B., & Loreto, V. (2019). General scores for accessibility and inequality measures in urban areas. Royal Society Open Science, 6(8).

[2] https://data.iledefrance.fr/explore/dataset/horaires-prevus-sur-les-lignes-de-transport-en-commun-dile-de-france-gtfs/information/

[3] Facebook data for good: Population Density Maps, https://dataforgood.fb.com/tools/population-density-maps/

[4] D. S. Johnson, J. K. Lenstra, and A. Rinnoiy Kan. "The complexity of the Network Design Problem". In: Networks (1978).

[5] A. Barreto, D. Silver, et al. "Successor features for transfer in reinforcement learning". In: NIPS (2017).

[6] H. Dai et al. "Learning combinatorial optimization algorithms over graphs". In: NIPS (2017).

[7] A. Mittal, A. Dhawan, S. Manchanda, S. Medya, S. Ranu, and A. Singh. "Learning heuristics over large graphs via deep reinforcement learning". In: arXiv (2019).