23/05/2024

Internship Log

Emanuele Mengoli

April 2024

Contents

1 Background

- NOMA - build a background. Read: [16, 27, 11, 7]
- Dynamic Cellular Network - Moving UEs. Read: [21]
- High-dimensional Bayesian Optimization - static case. Read: [22, 5]. Of [4] read Cap. 5,6. Read: [28]
- Mobility Models: see Section 2

2 Modelling Urban Mobility

2.1 State of Art Analysis

An extended list of mobility models for wireless networks is proposed in [6, 3, 14]. Camp et al., in [6] analyzed both group and entity mobility models for ad-hoc

adhoc networks are often based on peculiar technology, and not 4G or 5G networks, in order to stress-test novel protocol robustness. Bai et al., in [3] extended [6] by focusing on group mobility models. A Java-based open software implementation, which builds on [6], is presented in [23]. Similarly, [9] proposed a visual tool to observe taxi traffic mobility and trajectories based on the city of New York.

this could model a class of EVs. Another class could be pedestrians.

González et al., [10] provided a statistical measure of human routine dynamics, highlighting the intrinsic high degree of temporal and spatial regularity in human trajectories, with a significant probability of returning to a few highly frequented locations. The work suggests that pure random trajectories computed through Lévy flights and random walk models result in a high degree of inaccuracy.

Yan et al., [33] proposed a universal mobility model that leverages only population distribution information. A probability mapping of possible destinations is scored against the population density of the given area. The model relies on the assumption that the likelihood of a destination is inversely proportional to its population density, yet this belief ceases to be valid when considering the structure of European cities, whose architecture resembles concentric circles that progressively expand. In these scenarios, the population density tends to decay following a radial distribution, centred on the historical sector of the city, e.g. the area surrounding the cathedral. Furthermore, in general, museums and historic buildings are located within the historic centre, representing the largest area of interest for tourist activity, thus a significant destination of urban flow.

is it the case? or the inverse?

For a more realistic approach, trajectories have been estimated and studied through several means: road Access Points (APs) [30], mobile phone signaling data [13], Call Detail Records (CDR) [2], and historical data from CDR, GPS, and WiFi traces [15].

they collect the time at which UEs connect and disconnect to the BS?

Human mobility interpretation represents another stream of research, as highlighted by Loder et al., [18]. The authors proposed an interpretation framework for mobility behaviors to identify flow critical points that could lead to congestion.

Several works concern the reducibility of high-dimensional traffic datasets, such as [29], which leverages a Bayesian factorization framework to retrieve the inner spatial-temporal structure of a city in lower dimensions.

Finally, Uppoor et al., [31] made a step forward in vehicle mobility, following the lead of [17], by building a large-scale dataset descriptive of the Köln urban scenario.

Remarks: Most of the works presented leverage historical data from different sources and cities. This poses issues in terms of generalizability, since traffic trajectories are inherently shaped by the urban morphology and structure, as well as city dimensions and historical heritage of the considered urban scenario. European urban mobility scenarios are significantly different from their Asian or American counterparts due to the lower presence of skyscrapers and the medieval heritage of wall-protected concentric circles that have expanded radially over time. This is the case for cities like Paris, for example.

The Parisian mobility landscape has been modeled and analyzed in [2]. The proposed approach uses traffic data from CDR to model human behavior. Origin-Destination flows are estimated by employing the likelihood of the means of transportation, obtained from the given training set of trajectories. The full area of interest is divided into Voronoi sectors. Emanuele: I have requested both the data produced by their prediction model and used during the study from the authors (coming from a Mobile Operator). To date, I have not received any feedback.

Finally, two Python packages have been developed to include a wide range of random distributions for shaping mobility models: [20, 26]. The former allows building trajectories knowing points of interest within a given area, and the latter allows generating paths choosing a specific distribution.

Are many works using these packages? This could be an interesting option for our purpose

2.2 Dynamic Bayesian Optimization Framework

Our case scenario encompasses complex network dynamics, characterized by clusters of users populating the environment and subjected to spatial translation. In this term, a realistic scenario should consider the directional constraints posed by the territory morphology and the transport infrastructure, accompanied by its endemic mobility limits (i.e., velocity, direction, density...). The high degree of complexity would require access to city available data, allowing understanding of the Origin/Destination (O/D) flows, the population distribution, the traffic magnitude, and its distribution over time. Accessing the data would also allow trajectory interpretability, highlighting the main points of interest within the city and the purpose of travel, enabling estimates of users' trajectories and flows magnitude. With this consideration and introducing the assumption of a rational user, we can ease the probabilistic transition model as proposed in Section 2.3.

Nevertheless, according to the available literature on urban mobility for wireless networks, a first baseline to assess the Dynamic Bayesian Optimization (DBO) algorithm can be constructed using a random mobility model that mimics group dynamics. Therefore, the DBO can be backtested on two consecutive

meaning?

scenarios of ascending complexity.

The mobility model baseline can be constructed using a Reference Point Group Mobility Model (RPGM), that uses the Gauss–Markov Mobility Model (GMM) for entity granularity, to shape UE cluster trajectories in a pseudo-random fashion. Among the proposed models, the choice of combining these two is explainable by the following strengths:

references? more details

- Unlike pure Random Walk approaches, at the entity level, the GMM maintains temporal dependency in movement, enabling smooth changes thanks to the Markov property. This property attains realistic behaviors [3].

- RPGM is a well-known and widely used model for stress-testing network dynamics at the group level, as it maintains spatial dependency, i.e., correlated movement of clusters of UEs [12, 6, 3]. Regarding inter-group interaction, to simulate a significantly dense urban scenario, both in terms of population density and infrastructure, an inter-group *overlap* strategy can be taken into account. From a realistic point of view, it would model travelers using different means of transportation to cross a common area.

can the composition of clusters change over time?

We use the term hybrid-GMM to refer to the combination of these two algorithms. In practical terms, the hybrid-GMM is applied to the cluster centroid, triggering the displacement of the associated users. Upon reaching the destination, users are distributed uniformly within an ϵ distance from the new centroid position, therefore introducing an ulterior marginal degree of randomness within the model.

on their way to the destination, how are distributed EU?

2.3 Custom Probabilistic Mobility Model

The proposed probabilistic model is built as follows:
$X^{(j)} = \{s_t^{(j)}\}_{t=1}^{T^{(j)}}$ being the trajectory for the j-th user, such that $j \in \mathcal{N}$, where $\mathcal{N}$ is the set of users. Each user j has a total lifetime in the system indicated as $T^{(j)}$. Let $\{u_k\}_{k=1}^{K}$ be the set of means of transport. The area of interest is denoted as $\mathcal{A}$, where i denotes the i-th section, such that $\bigcup_{i\in\mathcal{A}} i = \mathcal{A}$ and $\bigcap_{i\in\mathcal{A}} i = \emptyset$, to ensure partition properties, $i \in \mathbb{N}$. For the purpose of our simulation, each i corresponds to a tile of the grid.

S is the coordinates of the user?

We can formulate transitions as follows:

$$\begin{aligned} s_{t+1}^{*(j)} &= \underset{i'\in\mathcal{A}_i}{\operatorname{argmax}}\, \mathbb{P}\left(s_{t+1}^{(j)} = i', s_t^{(j)} = i\right) \\ &= \underset{i'\in\mathcal{A}_i}{\operatorname{argmax}}\, \mathbb{P}\left(s_{t+1}^{(j)} = i' \mid s_t^{(j)} = i\right) \mathbb{P}\left(s_t^{(j)} = i\right) \end{aligned} \tag{1}$$

random draw at each iteration to select the next location of the user?

does this mean that user can only move from one section of interest to another?

shouldn't t=1 also be a random of i, e.g. $t^{(0)}$?

Such that $\mathcal{A}_i = \bigcup_{k=1}^{K} \mathcal{A}_{i,k}$ and $\mathcal{A}_{i,k} = \{i' \in \mathcal{A} \mid i' = i + \vec{v_k}\}$. Let $\vec{v_k}$ be average velocity associated with the k-th mean of transport, so that $\vec{v_k} = \vec{v_k}^{(t)}, t = 1, ..., \bigcup_{j \in \mathcal{N}} T^{(j)}$.

A logical approximation derives from considering a trajectory constrained by the choice of the mean of transport, both in terms of speed and reachable sector, due to infrastructure placement and ramifications. In fact, in realistic scenarios, travel flows are dependent on the type of road, rails, or more generally, traffic limitations in place. Thus, by considering displacement strictly dependent on the mode of transport, we could express the transition probability as follows: [1]

$$\begin{aligned} &\mathbb{P}\left(s_{t+1}^{(j)} = i' \mid s_t^{(j)} = i\right) \mathbb{P}\left(s_t^{(j)} = i\right) = \\ &\qquad \sum_{k=1}^{K} \mathbb{P}\left(u_{k,i',t+1}^{(j)} \mid u_{k,i,t}^{(j)}\right) \mathbb{P}\left(u_{k,i,t}^{(j)}\right) \end{aligned} \tag{2}$$

To the extent of our analysis, let the symbol " $\dashv$ " denote the operator that generalized the notion of *belonging* in non-mathematical sense, i.e. it indicates the means of transport passing through a tile or the user being on a given means of transport. Let $\mathbb{P}\left(u_{k,i,t}^{(j)}\right) = \mathbb{P}\left(j \dashv u_k, u_k \dashv i, s_t^{(j)} = i\right)$, such that:

$$\begin{aligned} &\mathbb{P}\left(j \dashv u_k, u_k \dashv i, s_t^{(j)} = i\right) = \\ &\quad \mathbb{P}\left(j \dashv u_k \mid u_k \dashv i, s_t^{(j)} = i\right) \mathbb{P}\left(u_k \dashv i, s_t^{(j)} = i\right) \end{aligned} \tag{3}$$

Due to independence, $\mathbb{P}\left(u_k \dashv i, s_t^{(j)} = i\right) = \mathbb{P}\left(u_k \dashv i\right) \mathbb{P}\left(s_t^{(j)} = i\right)$.

To simplify the model without losing interpretability, we could consider that on a given tile, for a given destination, users choose the means of transport based on the highest availability, which in our model embeds the notion of time to reach the destination and passing frequency. Thus, $\mathbb{P}\left(u_{k,i',t+1}^{(j)} \mid u_{k,i,t}^{(j)}\right) \approx \mathbb{P}\left(u_{k,i',t+1}^{(j)}\right)$. Furthermore, $\mathbb{P}\left(s_t^{(j)} = i\right)$ is computed considering the normalized average traffic flow density, Emanuele: i.e., people$/m^2$, of a given sector, obtained from the city's available data. Since it only provides a snapshot at a given time and not the entire real-time time series of the available traffic flow, we can approximate this probability as stationary in time and independent of the user. Therefore, $\mathbb{P}\left(s_t^{(j)} = i\right) = \mathbb{P}\left(s = i\right), t = 1, ..., \bigcup_{j \in \mathcal{N}} T^{(j)}$, which implies that $\mathbb{P}\left(u_{k,i,t}^{(j)}\right) = \mathbb{P}\left(u_{k,i}\right), t = 1, ..., \bigcup_{j \in \mathcal{N}} T^{(j)}$.

[1] Emanuele: Marginalizing by u_k implies that we do not need to take into account the user choice, as we naively assume the user takes the most available means of transportation. This defines an upper bound over the real case, where there is a higher degree of randomness in the user mobility choices.

Let γ denote state probability vector, such that $\gamma_i = \mathbb{P}(s = i)$. To ease notation, let $I = \mid \mathcal{A} \mid$, and let $U = \mathbb{P}(u_{k,i})_{k=1,\dots,K;\ i=1,\dots,I}$ be the $K \times I$ matrix that maps the joint distribution of urban transport means and user location on the grid in probabilistic terms. Also, let $W = \mathbb{P}(u_k \dashv i)_{k=1,\dots,K;\ i=1,\dots,I}$, and $\theta^{(j)} = \mathbb{P}(j \dashv u_k \mid \cdot\ )_{k=1,\dots,K} \sim \mathcal{U}(1, K)$, being $\mathcal{U}$ a uniform distribution. By assuming users to be IID on the transportation choice, $\theta^{(j)} = \theta, \forall j \in \mathcal{N}$.

We can then reduce our problem to an algebraic form as follows:

$$s_{t+1}^{*(j)} = \underset{i' \in \mathcal{A}_i}{\operatorname{argmax}}\, U_{i'}^\top U_i = \operatorname{argmax} U^\top U_i \tag{4}$$

$$U = \theta\gamma^\top \otimes W \tag{5}$$

Where " $\otimes$ " denotes the element-wise product. Therefore, the trajectory for the j-th user can be expressed as:

$$X^{(j)} = \left\{\operatorname{argmax} U^\top U_{s_{t-1}}\right\}_{t=1}^{T^{(j)}} \tag{6}$$

shouldn't this be associated to the time at which EV appears? t_0?

Where s_0 is the user's spawning tile.
The following databases provide insights regarding urban flow in the city of Paris: [19, 24, 34, 1].

3 Methodology

3.1 User Generation

If data are available, the initial pool of users is generated in the grid according to the distribution associated with the normalized traffic flow, i.e., γ. Alternatively, γ can be estimated through a normalized radial-basis density distribution. Let a given grid sector i have coordinates denoted as $\delta(i)$, and let the center coordinates be denoted as $\delta(c)$. Then,

$$\mathbb{P}(s = i) = \frac{e^{-\|\delta(i)-\delta(c)\|_2^2}}{\sum_{i \in \mathcal{A}} e^{-\|\delta(i)-\delta(c)\|_2^2}} \tag{7}$$

This follows the assumption that urban density decreases as the distance from the urban center increases. The notion of an urban center might not coincide with the geometric centroid of the grid, yet it represents the point in space where urban density reaches its peak. It can be estimated by analyzing the density of base stations within a neighborhood of a given radius or considering the traffic flow.

In our scenario, the center is defined as follows:

$$\delta(c) = \underset{\alpha \in \mathcal{A}}{\operatorname{argmax}} \lim_{r \to 0} \mid \mathcal{B}(\alpha, r) \mid$$
$$s.t.$$
$$\mid \mathcal{B}(\alpha, r) \mid > 1 \tag{8}$$
$$\alpha \in \mathbb{R}^2$$

How are BS position determined?

Let Z be the set of base stations within the grid, and let $\mathcal{B}(\alpha, r)$ be the ball encompassing the base stations that lie in its area. Let the operator "$| \cdot |$" extend the notion of set cardinality to the ball, indicating the number of base stations within the ball. Thus:

$$\mathcal{B}(\alpha, r) = \{z \in Z : \|\delta(z) - \alpha\|_2 \leq r\} \tag{9}$$

why M/M/1? FIFO?

Lastly, users' birth and death (B&D) process is modeled through an M/M/1 queue. Users are clustered using the K-means algorithm based on their initial spawning locations. Upon the birth of a new user, its placement follows the grid probability distribution (γ), and it is associated with the closest cluster, in terms of ℓ_2-norm. Cluster translation is triggered following an ϵ-greedy strategy.

do we set the number of clusters? K-means is not design to do this

what does this mean?

Do all UEs appear at t0? or do you recompute clusters after each UEs spawn?

3.2 Simulator Main Points

- Chosen episode frequency $\tau = 1s$: In LTE, one radio frame has $T_s = 10\,ms$ [8, 32], leading to 100 packets transmitted per episode.
- Mobility Models:
  1. Hybrid-GMM applied to UE clusters
  2. Probabilistic Mobility Model
- UE B&D Process:
  1. M/M/1 queue

How are they located?

3.3 Toy Test

The current implementation is available on Mengoli's GitHub - NOMA Simulator (`https://github.com/emanuelemengoli/NOMA-net-simulator`), see the file `test.ipynb`. The current simulation includes 734 base stations and 500 users, simulation time is set to 1000 episodes, with a total computational time of ≈ 24 min. Simulation parameter are set in `simulation_env.py`, and in the current version do not attain for realist values.

3.4 Next Steps

The next steps are articulated as follows:

- Develop a GIF-based visualization tool to observe how the simulation evolves over time on a 2D grid.
- *Prime* Visualization Tool: Develop a GIF-based realistic graphics tool, built on [25]. See and download Paris Map (`https://github.com/emanuelemengoli/NOMA-net-simulator/blob/main/output/map.html`) for the current implementation.
- Implement the Probabilistic Mobility Model and compute the required probability distributions from the dataset related to Parisian traffic.
- Generalize the Birth and Death process: Encompass a Renewal process to model UE birth. What do you mean?
- Include the DBO algorithm.

References

[1] R. autonome des transports parisiens (RATP). Ratp data platform. `https://data.ratp.fr/explore/?sort=modified`.

[2] D. Bachir, G. Khodabandelou, V. Gauthier, M. El Yacoubi, and J. Puchinger. Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 101:254–275, 2019.

[3] F. Bai and A. Helmy. A survey of mobility models. *Wireless Adhoc Networks. University of Southern California, USA*, 206:147, 2004.

[4] A. Bardou. *Online Learning for the Black-Box Optimization of Wireless Networks*. PhD thesis, École Normale Supérieure de Lyon, 2023.

[5] A. Bardou, P. Thiran, and T. Begin. Relaxing the additivity constraints in decentralized no-regret high-dimensional bayesian optimization. *arXiv preprint arXiv:2305.19838*, 2023.

[6] T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless communications and mobile computing*, 2(5):483–502, 2002.