Mining Constrained Regions of Interest: An Optimization Approach⋆

Alexandre Dubray, Guillaume Derval, Siegfried Nijssen, and Pierre Schaus

UCLouvain - ICTEAM/INGI, Louvain-la-Neuve, Belgium
{first}.{second}@uclouvain.be

The number of devices producing spatiotemporal data is increasing and so the data that need to be processed by the applications. It is important to have generic and flexible tools to be able to summarize and analyze this data efficiently. One such tool is the discovery of Regions of Interest (ROI) or densely visited regions. The discovery of ROI can also be instrumental as a preprocessing step to rewrite the trajectories as a sequence of interesting places instead of a sequence of GPS data point, which are easier to process and match by frequent sequence analysis tools [1].

There exists different algorithms to extract ROIs, depending on the definition of density we choose. In our paper we are interested in grid-based approaches which first divide the map into a grid and assign a density value to each cell. The density of a cell is the number of trajectories crossing it. A cell is dense when its density value is above a threshold. Then the ROIs are formed by aggregating, in some ways, the dense cells, similarly to a clustering algorithm. Multiple approaches exist to aggregate the dense cells in ROIs, generally based on a greedy expansion or clustering. While these methods give good results, they do not easily accept new constraints such as various types of shapes, and application dependent intra- or inter-ROI constraints.

To address these weaknesses, we propose a new approach to extract *Constrained Regions of Interest*, which is illustrated in Figure 1. This two-step optimization process allows to impose constraints on individual ROI (*intra-ROI* constraints) and between the ROIs (*inter-ROI* constraints).

The first step is to generate a set of candidates ROI (e.g. rectangles, circles, or any non-parametric shapes) that respects the *intra-ROI* constraints such as "Every ROI must contain at least one cultural point of interest".

Once we have this set of candidates, we will select the ROIs to return to the user, while ensuring they respect *inter ROIs* constraints. The idea is to consider that a set of ROI is a classifier that indicates whether a cell on the grid is dense or not. Such classifier would make some errors (a dense cell not covered by a ROI or a non-dense cell covered by a ROI) and we are interested in a set of K ROIs that minimizes the number of errors.

We propose an efficient *Integer Linear Program* (ILP) model to solve this problem with one binary decision variable per candidate with the constraint that two ROIs cannot overlap. We use the *Minimum Description Length* principle to automatically detect the appropriate K. During this phase we can add

⋆ This paper is a summary of a paper with the same name published at Discovery Science 2020 https://doi.org/10.1007/978-3-030-61527-7_41.

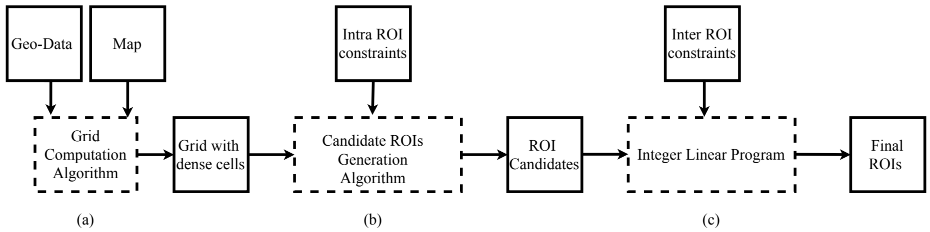

Fig. 1: Process of our approach: a) Creation of the density grid from the trajectories; b) Generation of candidates shape on the grid; c) Selection of the ROIs from the set of candidates.

constraints between the ROI as long as they can be translated into a linear constraint. For example, the constraint "If we select this ROI, then that ROI must also be selected" can be expressed using classical boolean logic constraints on the decision variables.

We compared our method to *PopularRegions* [2], a method specifically designed to extract ROIs, and OPTICS [3] when clustering the dense cells. We show that our method is slower than the others but, as the number of candidates decreases, our run time becomes like OPTICS. Moreover we obtain a lower description length with a better balance between the errors and the number of ROIs. This allows our method to be more flexible and have a better generalization. Finally, we show that when adding up to 40% of noise in the data, our method is more stable than *PopularRegion* and OPTICS. We refer to the complete article [4] for further information on the approach.

References

1. John OR Aoga, Tias Guns, and Pierre Schaus. Mining time-constrained sequential patterns with constraint programming. *Constraints*, 22(4):548–570, 2017.
2. Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. Trajectory pattern mining. In *SIGKDD*, 2007.
3. Mihael Ankerst, Markus M Breunig, Hans-Peter Kriegel, and Jörg Sander. Optics: ordering points to identify the clustering structure. *ACM Sigmod record*, 1999.
4. Alexandre Dubray, Guillaume Derval, Siegfried Nijssen, and Pierre Schaus. Mining constrained regions of interest: An optimization approach. In *International Conference on Discovery Science*, pages 630–644. Springer, 2020.