View Reviews

Paper ID

3

Paper Title

Colossal Trajectory Mining: Semantic Co-movement Pattern Mining

Track Name

SEBD2024

Reviewer #1

Questions

1. Main contribution and comments to the authors

the paper summarises a prior publication:
M. Francia, E. Gallinucci, M. Golfarelli, Colossal trajectory mining: A unifying approach to mine behavioral mobility patterns, Expert Syst. Appl. 238 (2024) 122055. doi: 10.1016/ J.ESWA.2023.122055.

it presents a unified framework for the extraction of heterogeneous mobility patterns, which scales over large volumes of trajectories with the assumption that these lay within a limited combination of time and space

2. Strengths (bullet points)

S1 very recently previously well-published work
S2 implementation (Spark) exploits parallelilsm -- but it's not not really explained in detail

3. Weaknesses (bullet points)

W1 the paper is effectively a teaser: the arguably more interesting part, nanely the actual algorithm, is glossed over in sec 3
W2 unclear where the Milan data comes from: is that consented from citizens? where is the data from?
W3 this may be controversial? "The attractiveness of a neighborhood is defined as the percentage of co-movement patterns passing through that neighborhood. " I can imagine neighboroods that are desirable because they are very quiet in terms of passing traffic...

4. Recommendation

Weak accept

5. Relevance

Relevant

Reviewer #2

Questions

1. Main contribution and comments to the authors

This discussion paper presents a framework for general clustering of massive trajectories dataset. it has been extracted by a cited journal paper that provdes more details.

For example, recent papers at SIGSPATIAL should be cited and very brifley compared. Why your tesseletion is different and which are the advantages of your tesseletion compared with other approaches (I am talking only about tesselation and trajectory summarization, not the clustering):

Ali Faraji, Jing Li, Gian Alix, Mahmoud Alsaeed, Nina Yanin, Amirhossein Nadiri, and Manos Papagelis. 2023. Point2Hex: Higher-order Mobility Flow Data and Resources. In Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23). Association for Computing Machinery, New York, NY, USA, Article 69, 1–4. https://doi.org/10.1145/3589132.3625619

Chiara Pugliese, Francesco Lettich, Fabio Pinelli, and Chiara Renso. 2023. Summarizing Trajectories Using Semantically Enriched Geographical Context. In Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23). Association for Computing Machinery, New York, NY, USA, Article 44, 1–10. https://doi.org/10.1145/3589132.3625587

Maria Luisa Damiani, Fatima Hachem, Christian Quadri, Matteo Rossini, Sabrina Gaito: On Location Relevance and Diversity in Human Mobility Data. ACM Trans. Spatial Algorithms Syst. 7(2): 7:1-7:38 (2021)

2. Strengths (bullet points)

the proposed problem is indeed of interest when analysing big dataset of movement data
the approach is tested in 3 datasets (2 publicly avaiavble) showing good results
paper is well written and it can be of interest for the SEBD audience

3. Weaknesses (bullet points)

I think there are no critical weaknesses but I think that some part of the literature on trajectories represented with the tesellation is not included and should be added.

4. Recommendation

Accept

5. Relevance

Relevant

Reviewer #3

Questions

1. Main contribution and comments to the authors

The paper is a concise version of ref. [12], and mainly concentrates on motivating aspects and on the model underlying trajectory mining. The idea is interesting, even if it is unclear to which extent the approach can be applied to more fine-grained cases. However, the paper has some weaknesses that should somehow be fixed.

The term "colossal" was used in [12], yet it should be credited to:
Feida Zhu, Xifeng Yan, Jiawei Han, Philip S. Yu, Hong Cheng:
Mining Colossal Frequent Patterns by Core Pattern Fusion. ICDE 2007.
where the intuitive definition is:
"longer sequences are usually of more significant meaning than shorter ones in bioinfomatics. We call these large patterns colossal patterns, as distinguished from the patterns with large support set."
Is this also true in the CTM model?

The definition of tessellation is a central one, yet Def. 2 seems to have some problems:
- The notion of adjacency is given for each feature independently of the others. Thus, two tiles A and B could be adjacent on feature f1 but not on f2. It seems that a notion of "global" adjacency is missing here. Please clarify.
- The very definition of adjacency on feature f is flawed and needs to be fixed: if dist is a metric the triangular inequality always holds, thus any two tiles are adjacent! On a linear domain (e.g., time) this is always true.
- The intuitive definition of geodesic distance (= length of shortest path between two tiles in the tessellation) is somehow ambiguous. Consider a regular grid, as in Fig. 2 (the simplest case for the spatial feature), and tiles A1 and B2: is their distance 1 (moving on the diagonal) or 2 (i.e., Manhattan distance)?

The problem is that adjacency is defined in terms of distance, and distance is (implicitly) defined in terms of adjacency (is "neighboring tiles" the same as adjacent tiles?)

Def. 3 is unclear: How can two tiles be disconnected? A tessellation is a partitioning of the feature space F. With a single feature (e.g., 2D space) absence of a path between two tiles implies that F is partitioned into (at least 2) disjoint parts, which can therefore be separately analyzed (e.g., the 2D space includes Milano and Cagliari).
Again, the def. is given for a single feature

The only part in which notation and defs of Section 3.1 are used is in Table 2, which is not referenced at all in the text and no comments are given. Being the emphasis of this paper in modeling aspects this is quite surprising. The same lack of comments plagues Figure 3.

At the end of page 4 a sentence should be fixed:
"we need to identify the itemsets contained in a large number of transactions. As for FIM, our goal is identifying the itemsets
contained in a large number of transactions."

2. Strengths (bullet points)

- A model for trajectory mining
- Experiments on real datasets

3. Weaknesses (bullet points)

- The description of the model needs to be made more precise and consistent
- Unclear description of the algorithm
- Table 2 needs comments

4. Recommendation

Weak reject

5. Relevance

Relevant