10th Brazilian Conference on Intelligent Systems (BRACIS)

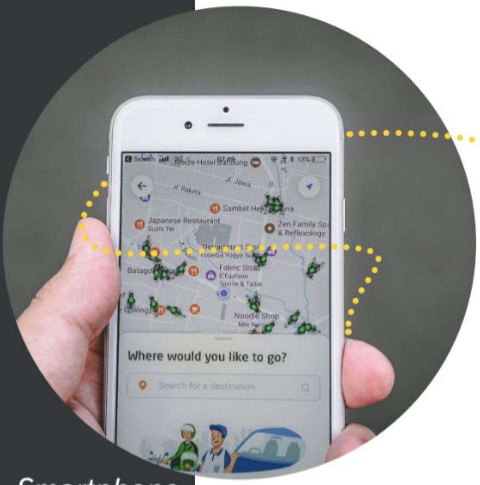
Fast Movelet Extraction and Dimensionality Reduction for Robust Multiple Aspect Trajectory Classification

Tarlis Tortelli Portela
Camila Leite da Silva
Jonata Tyska Carvalho
Vania Bogorny

CAPES

This work has been partially supported by the Brazilian agencies CAPES, CNPQ and FAPESC.

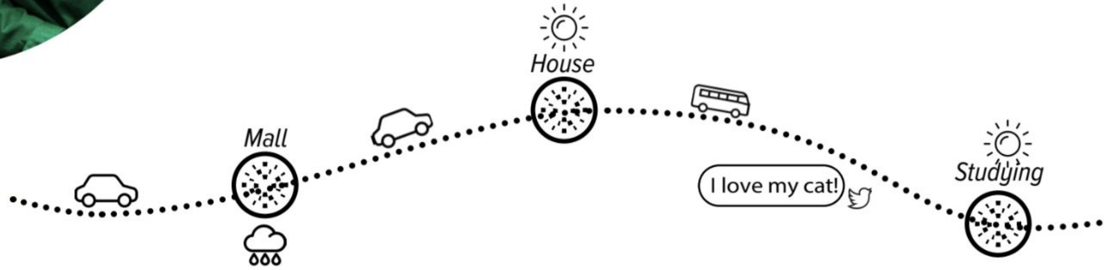
INTRODUCTION
Context



Smartphone Application

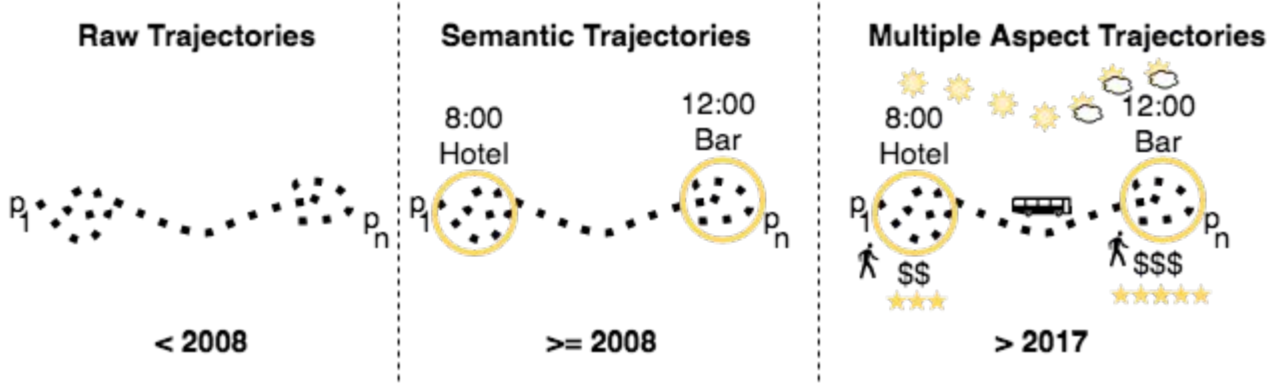
Popularization and price reduction of mobile devices

- Large volumes of mobility data;
- **Big Data.**



INTRODUCTION

Trajectory



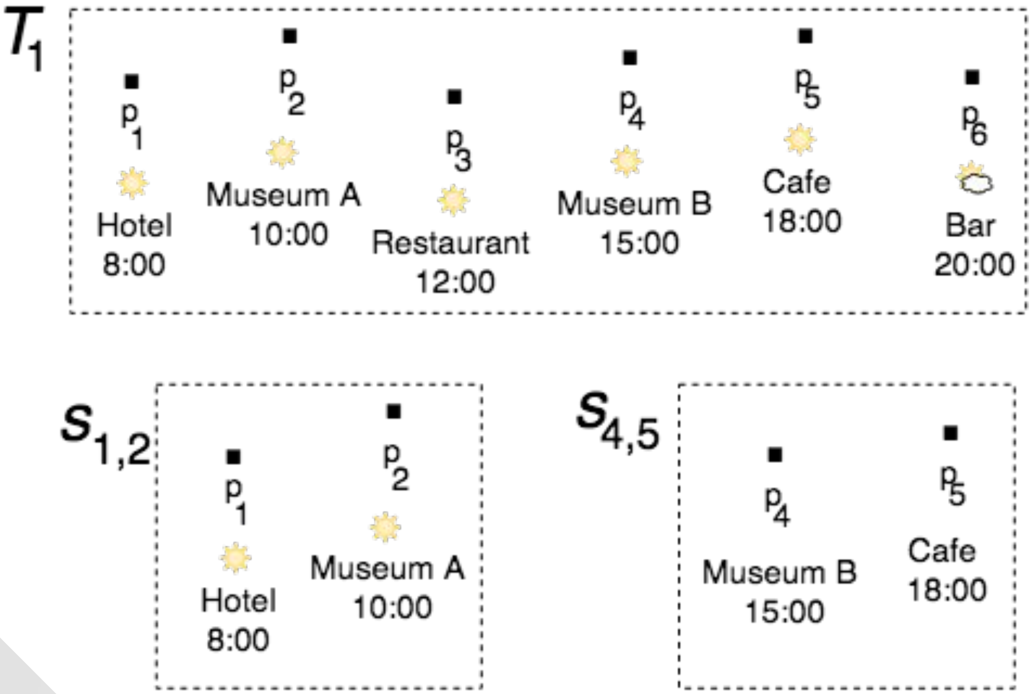
→ A **Multiple Aspect Trajectory** T_i is a sequence of *m* elements $T_i = \langle e_1, e_2, ..., e_m \rangle$, where each element is characterized by a set of *l* dimensions $D = \{ d_1, d_2, ..., d_l \}$, also called aspects.

→ Multiple and heterogeneous dimensions.

[Ferrero et al., 2016; Mello et al., 2019]

INTRODUCTION:

Trajectory and Subtrajectory



But,
What is a movelet?

Inspired by time series shapelet [Ye, L.; Keogh, E., 2011] a movelet is a subtrajectory that used by a classifier, better discriminate a class

INTRODUCTION

Trajectory Classification

- Trajectory data mining is important for discovering interesting knowledge and behavior about different objects as people, animals, vehicles, weather condition;

- An important data mining technique is classification:

Trajectory classification is the task of discovering the class label of a moving object based on its trajectories (Lee et al.2008).

INTRODUCTION

Motivation

Applications of Trajectory Classification

a. Transportation mean classification;
b. The strength level of a hurricane / Natural disaster prediction;
c. The type of a vessel;
d. Animal categories
e. The moving object, owner of the trajectory.

RELATED WORKS
Trajectory Classification

1st Semantic Classification
(TRAGOPOULOU; VARLAMIS;EIRINAKI, **2014**)
(VARLAMIS, **2015**)

Transportation Mean
(XIAO et al., **2017**)
(ETEMAD; SOARES JÚNIOR; MATWIN, **2018**)
(DABIRI; HEASLIP, **2018**)

General
(SANTOS; JR; ALVARES, **2011**)

***Survey* ***
(LEITE DA SILVA; MAY PETRY; BOGORNY, **2019**)

POI-F
(VICENZI et al., **2020**)

***Deals with multiple dimensions**

Movelets
(FERRERO et al., **2018**)

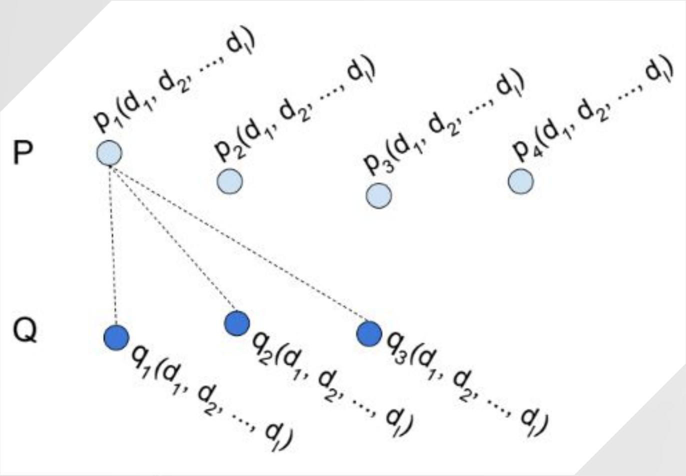
MASTERMovelets
(FERRERO et al., **2020**)

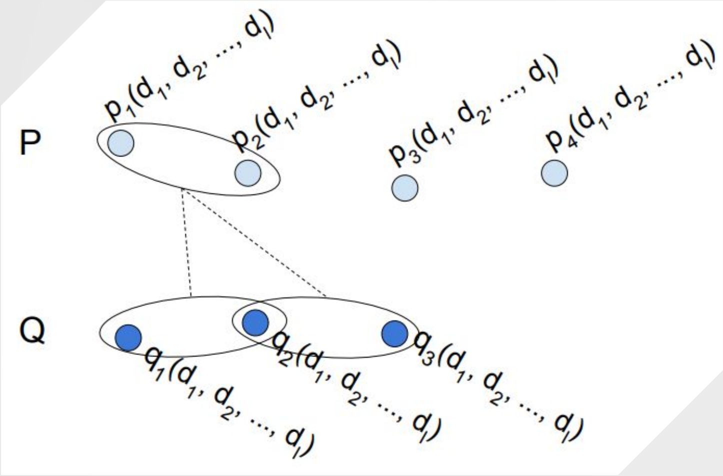
MARC
(MAY PETRY et al., **2020**)

RELATED WORKS

Movelets and MASTERMovelets

- Parameter free;
- Analyze every possible subtrajectory and computes the distance of all subtrajectories of the same size in the dataset.





RELATED WORKS

MASTERMovelets Method Overview

1. Computes element distances (point-to-point)
2. Extracts all subtrajectories with any subset of dimensions (movelet candidates)
3. For each candidate, compares to all trajectories:
    a. Find the **best alignment** (in each position of a trajectory)
    b. Evaluates the F-Score (split the classes)
4. Select the best movelets

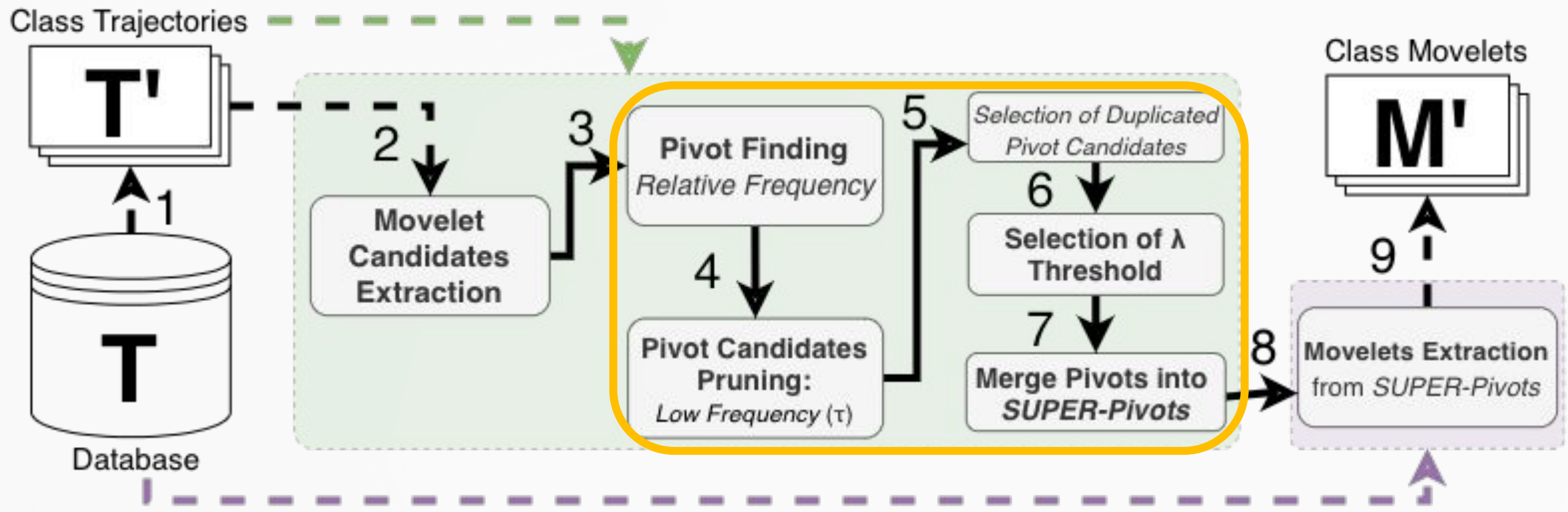
PROBLEM DEFINITION

- The problem of trajectory classification relies on finding the best ***trajectory*** or ***subtrajectory*** features to use as input to a classifier [Ferrero te al., 2020];
- Related works do not propose new classifiers (RF, NN, DT)
- So far, movelets [Ferrero te al., 2019] has been one of the best approaches:
  - highest accuracy
  - general problems
  - Interpretable patterns;

- **Problem**:
  - Movelets extraction is very time consuming.

PROPOSAL

SUPERMovelets



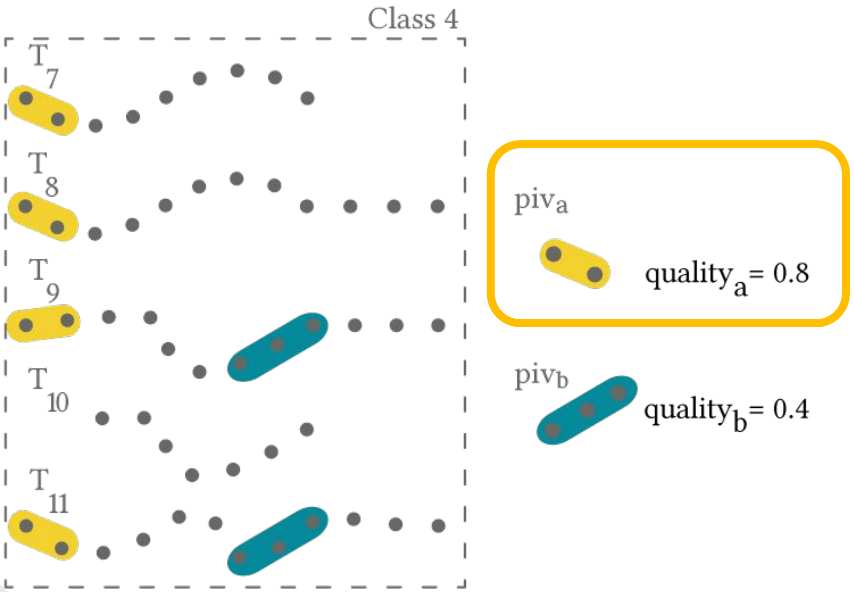
Fast Movelet Extraction and Dimensionality Reduction for Robust Multiple Aspect Trajectory Classification

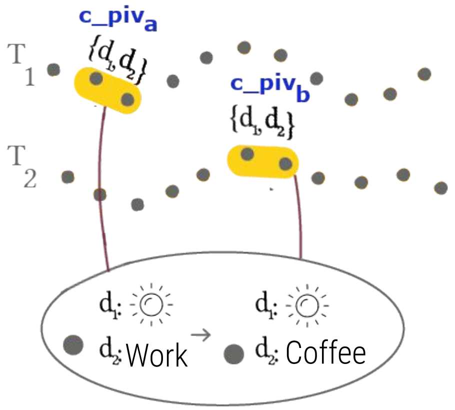
PROPOSAL

SUPERMovelets

Search space is reduced by the SUPER-Pivots:

- **Quality proportion** of the pivot candidates;
- **Redundant** pivot candidates (same size and trajectory dimensions).





EXPERIMENTAL RESULTS

Highlights

- **SUPERMovelets movelet extraction:**
  - Movelet extraction 50-94% faster than MASTERMovelets;
- **SUPERMovelets accuracy:**
  - Same accuracy as MASTERMovelets (less than 1% difference);
  - Generates significantly less movelets (65-93% reduction);
  - Faster to build classification models.

REFERENCES

- ETEMAD, M.; Soares Júnior, A.; MATWIN, S. Predicting transportation modes of GPS trajectories using feature engineering and noise removal. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v. 10832 LNAI, n. ii, p. 259–264, 2018. ISSN 16113349.
- FERRERO, C. A.; ALVARES, L. O.; BOGORNY, V. Multiple aspect trajectory data analysis: Research challenges and opportunities. Proceedings of the Brazilian Symposiumon GeoInformatics, v. 2016-Novem, p. 56–67, 2016. ISSN 21794847.
- FERRERO, C. A. et al. MOVELETS: Exploring relevant subtrajectories for robust trajectory classification. Proceedings of the 33rd Annual ACM Symposium on Applied Computing, Association for Computing Machinery, New York, NY, USA, n. April, p. 849–856, 2018.
- FERRERO, C. A. et al. MasterMovelets: discovering heterogeneous movelets for multiple aspect trajectory classification.Data Min. Knowl. Discov., v. 34, n. 3, p. 652–680, 2020.
- LEE, J., HAN, J., Li, X., GONZALEZ, H.: Traclass: trajectory classification using hierarchical region-based and trajectory-based clustering. Proc. VLDB Endow.1(1),1081–1094 (2008
- LEITE DA SILVA, C.; MAY PETRY, L.; BOGORNY, V. A Survey and Comparison of Trajectory Classification Methods. 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), p. 788–793, 2019.
- MAY PETRY, L. et al. MARC: a robust method for multiple-aspect trajectory classification viaspace, time, and semantic embeddings. International Journal of Geographical Information Science, 2020. ISSN 13623087.

REFERENCES

- MELLO, R. d. S. et al. MASTER: A multiple aspect view on trajectories. Transactions in GIS, 2019. ISSN 14679671.
- MUEEN, A.; KEOGH, E.; YOUNG, N. Logical-shapelets: An expressive primitive for time series classification. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p. 1154–1162, 2011.
- RAKTHANMANON, T.; KEOGH, E. Fast shapelets: A scalable algorithm for discovering time series shapelets. Proceedings of the 2013 SIAM International Conference on Data Mining, SDM 2013, p. 668–676, 2013.
- SPACCAPIETRA, S. et al. A conceptual view on trajectories. Data and Knowledge Engineering, Elsevier, v. 65, n. 1, p. 126–146, 2008. ISSN 0169023X.
- TRAGOPOULOU, S.; VARLAMIS, I.; EIRINAKI, M. Classification of movement data concerning user's activity recognition via mobile phones. ACM International Conference Proceeding Series, 2014.
- VARLAMIS, I. Evolutionary data sampling for user movement classification.2015 IEEE Congress on Evolutionary Computation, CEC 2015 - Proceedings, IEEE, p. 730–737,2015.
- XIAO, Z. et al. Identifying different transportation modes from trajectory data using tree-based ensemble classifiers. ISPRS International Journal of Geo-Information, v. 6, n. 2, 2017.ISSN 22209964.
- YE, L.; KEOGH, E. Time series shapelets: A novel technique that allows accurate, interpretable and fast classification. Data Mining and Knowledge Discovery, v. 22, n. 1-2, p.149–182, 2011. ISSN 13845810.
- ZHANG, Z. et al. Discriminative extraction of features from time series. Neurocomputing, v. 275, p. 2317–2328, 2018. ISSN 18728286.
- ZUO, J.; ZEITOUNI, K.; TAHER, Y. SE4TeC: A scalable engine for efficient and expressive time series classification. CEUR Workshop Proceedings, v. 2343, p. 8–12, 2018. ISSN 16130073.

10th Brazilian Conference on Intelligent Systems (BRACIS)

THANK YOU

Fast Movelet Extraction and Dimensionality Reduction for Robust Multiple Aspect Trajectory Classification

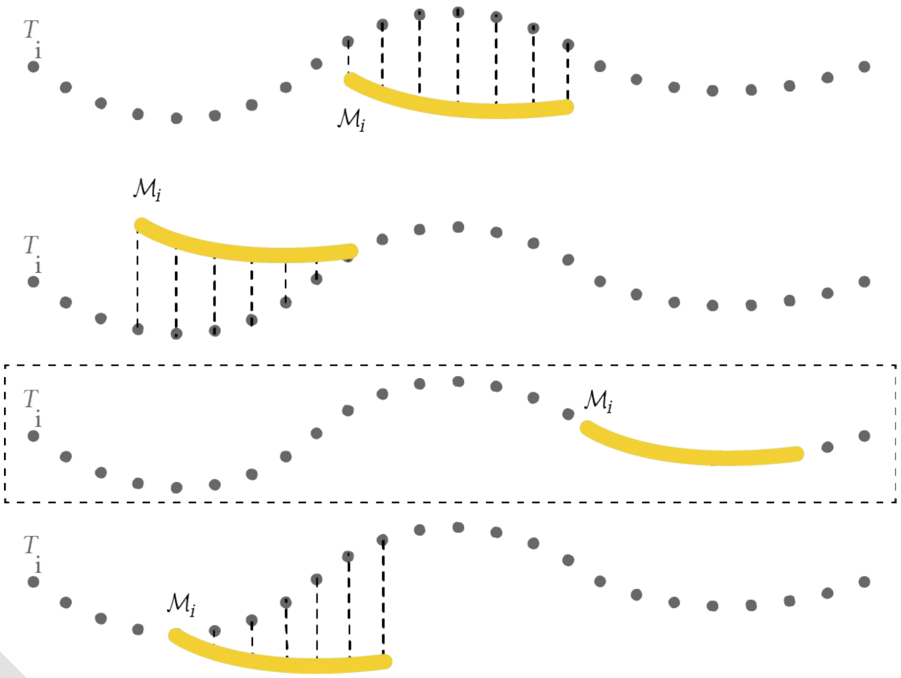
Tarlis Tortelli Portela
Camila Leite da Silva
Jonata Tyska Carvalho
Vania Bogorny

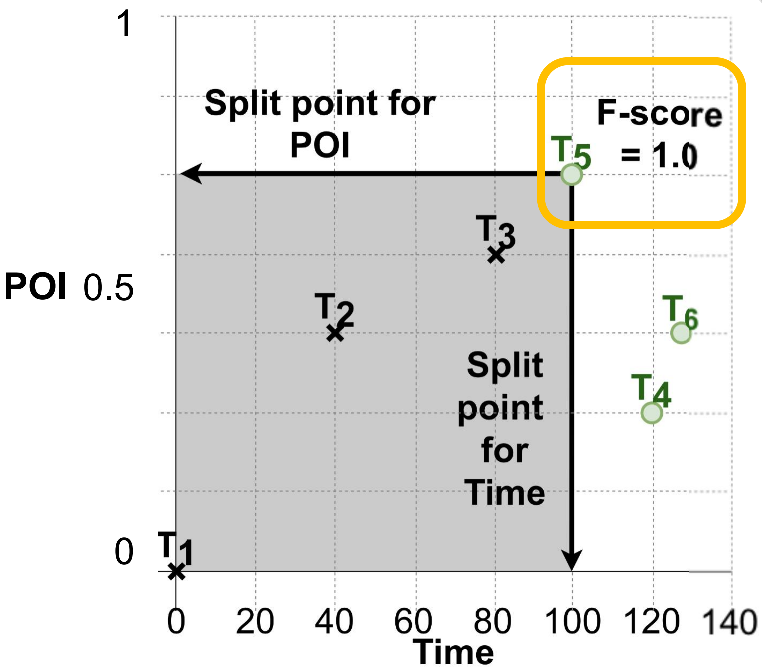
CAPES

This work has been partially supported by the Brazilian agencies CAPES, CNPQ and FAPESC.

BASIC CONCEPTS

Finding the Best Alignment





BASIC CONCEPTS

MASTERMovelets Complexity

- **n** → the number of trajectories;
- **m** → the length of the longest trajectory, and;
- **l** → the number of dimensions of the dataset

Space: it stores at most n × m candidates for all trajectories.

$O(n \times m^2 \times l)$ → *Matrix of Distances* *

Running Time: the overall time complexity is

$O(n^3 \times m^3 \log m \times 2^l)$

EXPERIMENTAL RESULTS

Scalability: Number of Dimensions

All experiments: faster as dimensions are added

EXPERIMENTAL RESULTS

Scalability:
Number of Points

All experiments:
faster as points are added

EXPERIMENTAL RESULTS

Scalability: Number of Trajectories

All experiments: faster as trajectories are added