UNIVERSITÀ DI PISA
CONSIGLIO NAZIONALE DELLE RICERCHE
UNIVERSIDADE FEDERAL DE SANTA CATARINA

Mauriana Pesaresi
Seminar Series 2021/2022

Movelet-based Classification of Multiple Aspect Trajectories

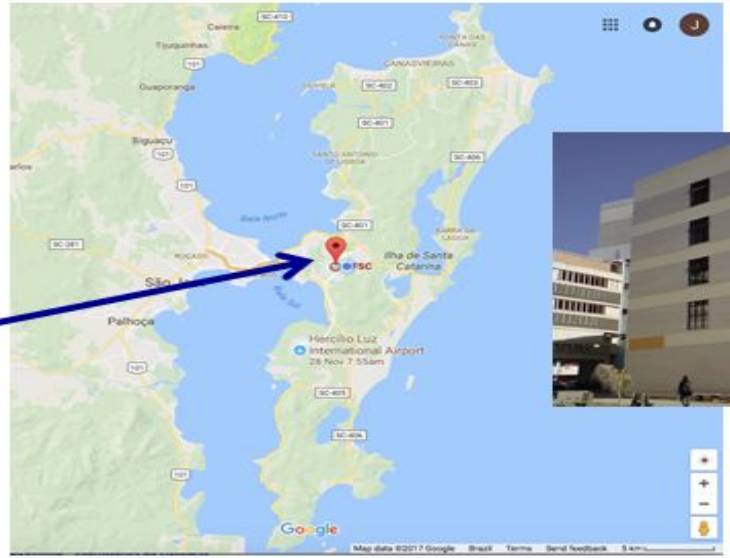
Tarlis Tortelli Portela

Advisors: Anna Bernasconi, Ph.D.
Chiara Renso, Ph.D.
Vania Bogorny, Ph.D.

Spatial Location

Universidade Federal de Santa Catarina

Campus Florianópolis



Building: Dept. de Informática e Estatística (INE)

UFSC: Universidade Federal de Santa Catarina

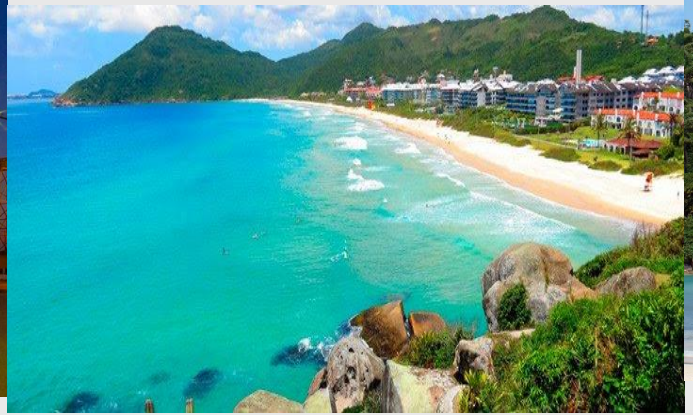
UFSC main campus

Founded in 1960, UFSC offers all Programs for free (no tuition)

- 107 Undergraduate Programs
- 65 Master Programs
- 56 PhD Programs

- 30.000 undergraduate students
- 15.000 graduate students

5th Best University in Brazil
12th Best in Latin America



AGENDA

01 **INTRODUCTION**
Context
Trajectory Classification
Motivation

02 **BASIC CONCEPTS**
Trajectory
Movelets
Complexity

03 **PROBLEM DEFINITION**
Research Questions
Related Works

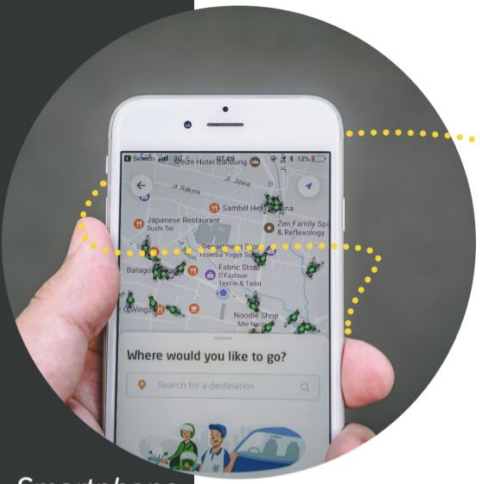
04 **MOVELET EXTRACTION**
Concepts
Method Complexity

05 **METHODS**
SUPERMovelets
HiPerMovelets

06 **SUMMARY**
Current State
Open Issues
References

INTRODUCTION

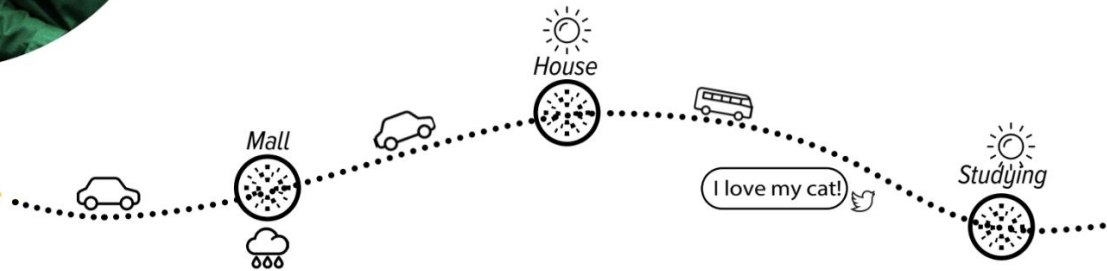
Context



Smartphone Application

Popularization and price reduction of mobile devices

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



INTRODUCTION

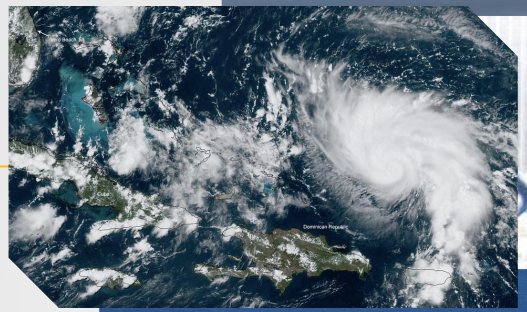
Trajectory Classification

- 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).

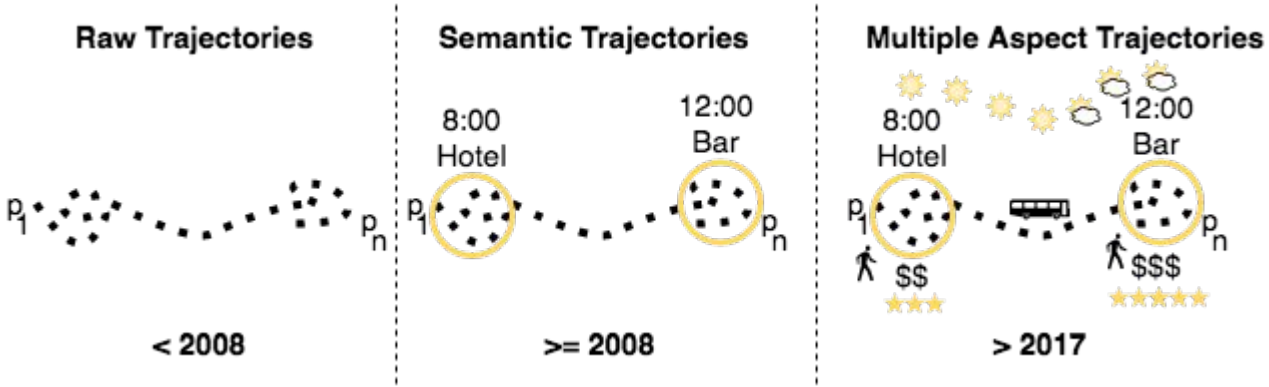
INTRODUCTION
Motivation



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.

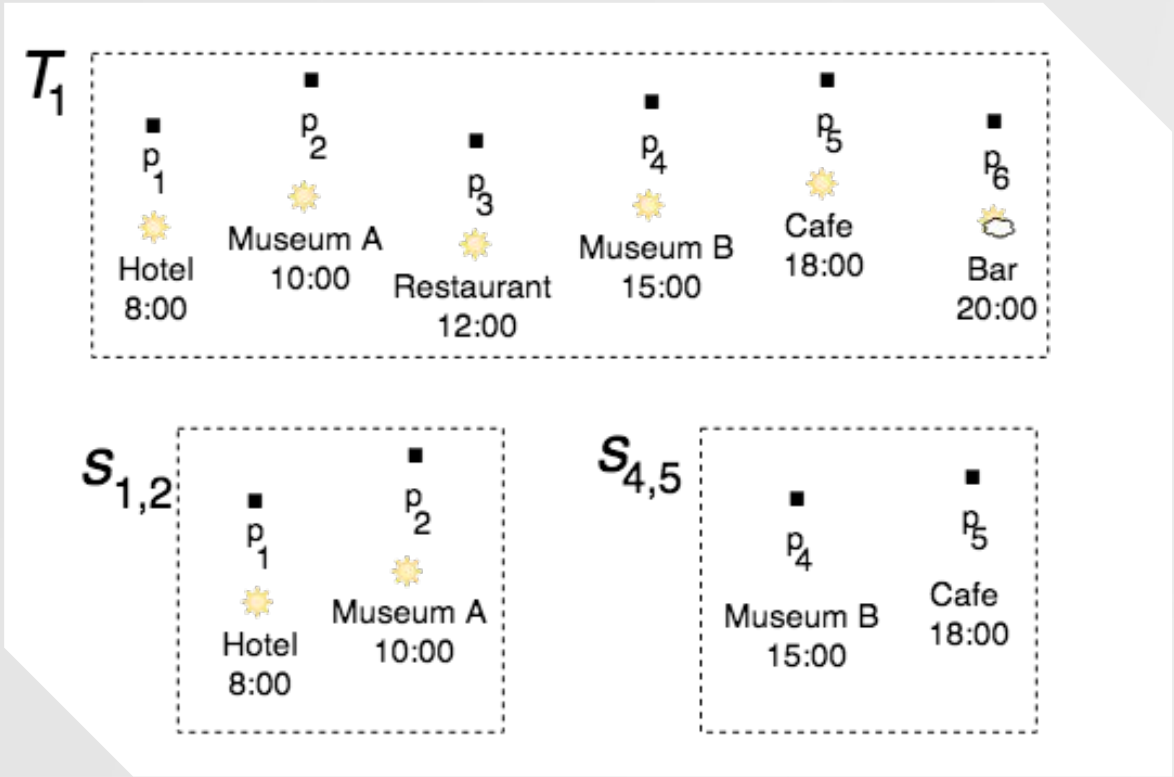
BASIC CONCEPTS:
Trajectory



→ 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.

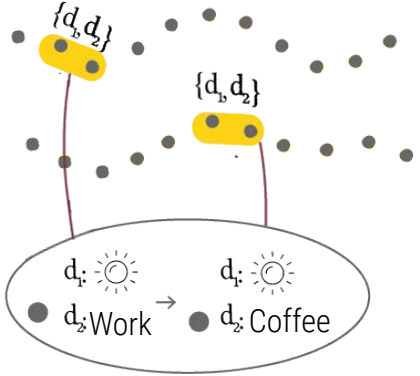
→ Multiple and heterogeneous dimensions.
[Ferrero et al., 2016; Mello et al., 2019]

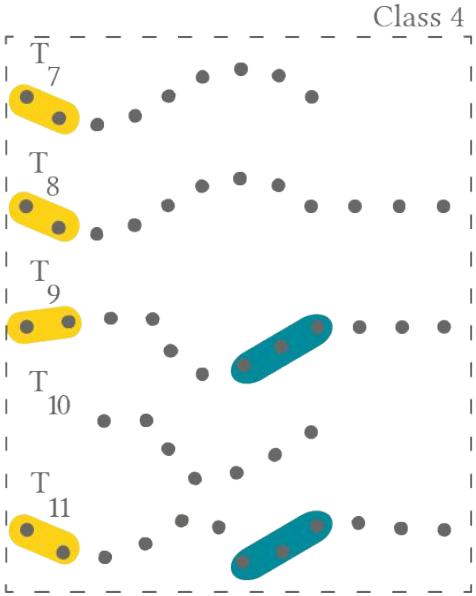
BASIC CONCEPTS:
Subtrajectory



Basic Concepts: 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:





Movelet:

(a)

$quality_a = 0.8$

(b)

$quality_b = 0.4$

PROBLEM DEFINITION

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

Trajectory Classification Related Works

1st Semantic Classification

(TRAGOPOULOU; VARLAMIS;EIRINAKI, **2014**)
(VARLAMIS, **2015**)

Transportation Mean

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

General

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

POI-F

(VICENZI et al., **2020**)

Survey *

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

MARC

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

Movelets

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

MASTERMovelets

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

SUPERMovelets

(PORTELA et al., **2021**)

HiPerMovelets

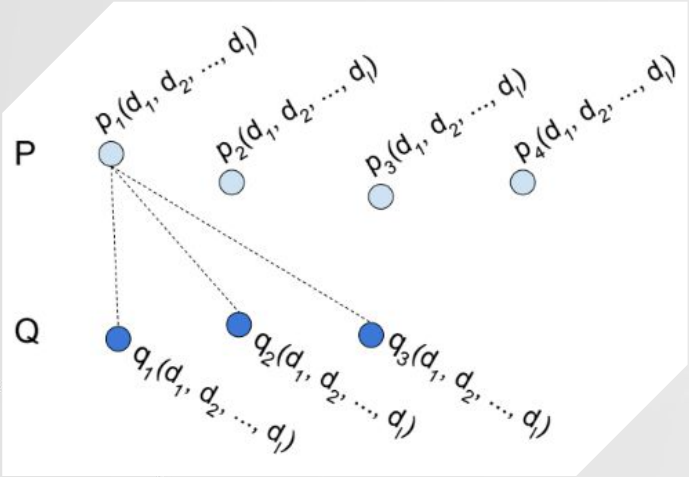
(PORTELA et al., **2022**)

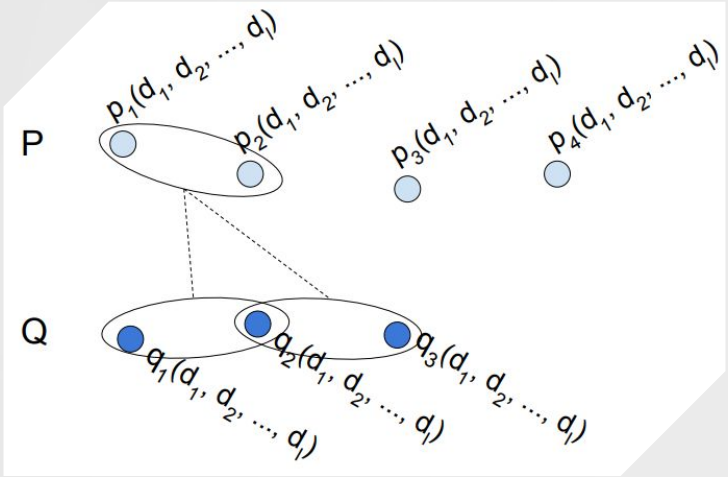
***Deals with multiple dimensions**

MOVELETS EXTRACTION

MASTERMovelets

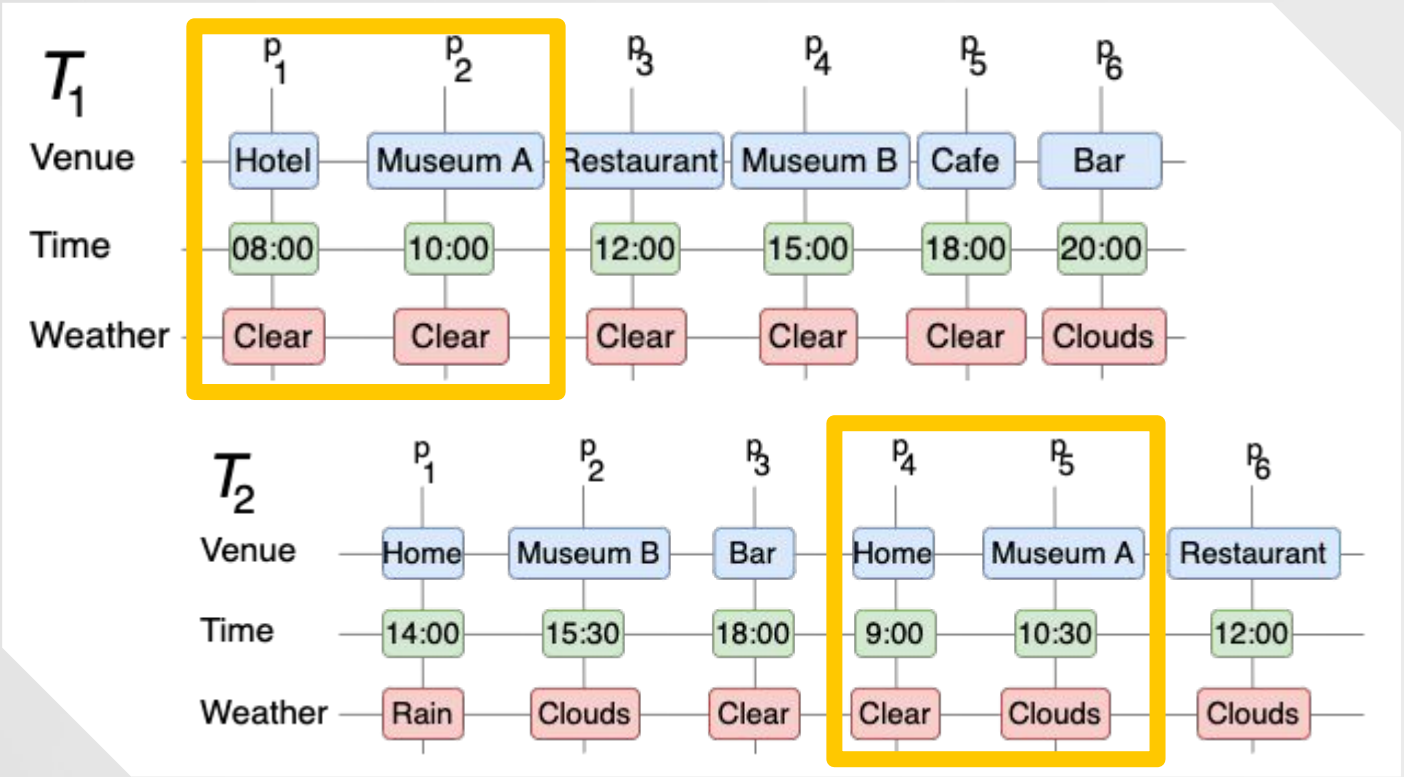
- Parameter free;
- Analyze every possible subtrajectory and computes the distance of all subtrajectories of the same size in the dataset.



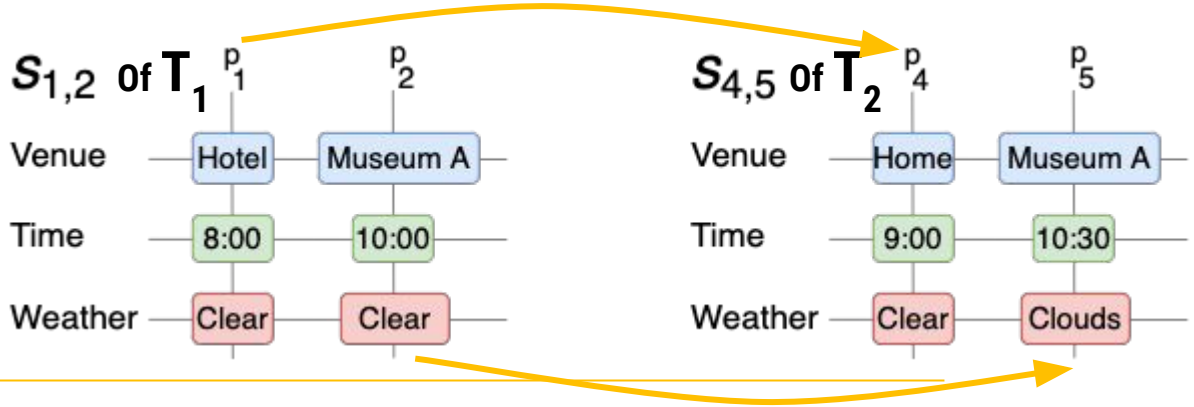


RESEARCH PROBLEM:

How to efficiently extract movelets?



Basic Concepts: *Subtrajectory Distances*



Element Distances	
1 (Hotel <> Home)	0 (Museum A <> Museum A)
60 (8:00<> 9:00)	30 (10:00 <> 10:30)
0 (Clear <> Clear)	0 (Clear <> Clouds)

$\Rightarrow$

Subtrajectory Distances (V)	
v_1	0.5 ((1 + 0) / 2)
v_2	45 ((60 + 30) / 2)
v_3	0.5 ((0 + 1) / 2)

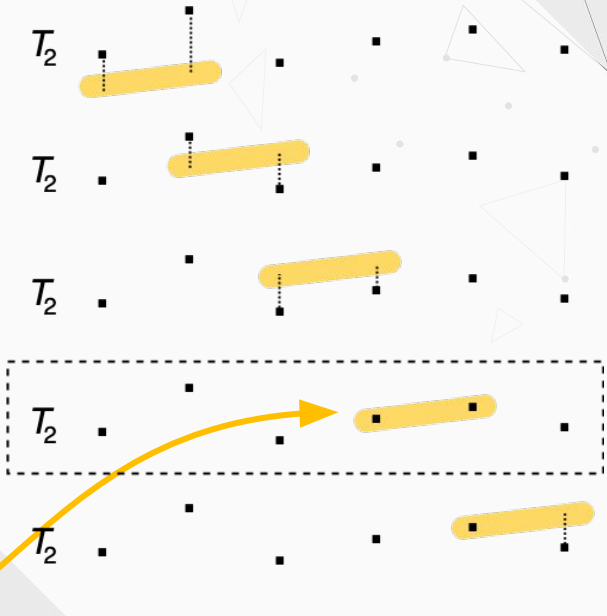
MOVELETS EXTRACTION

Ranking and Best Alignment

- Ranking distances

Distance	Starting Position				
	p_1	p_2	p_3	p_4	p_5
Venue	1	1	1	0.5	1
Time	345	465	330	45	135
Weather	1	0.5	0	0.5	1
Vector	$\mathbf{V}_1$	$\mathbf{V}_2$	$\mathbf{V}_3$	$\mathbf{V}_4$	$\mathbf{V}_5$

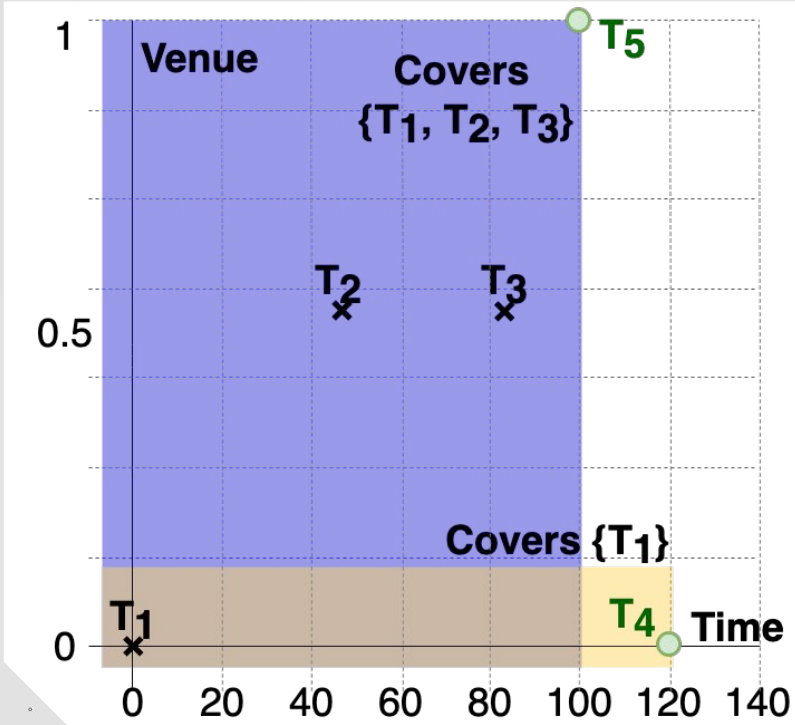
Rankings	Starting Position				
	p_1	p_2	p_3	p_4	p_5
Venue	2	2	2	1	2
Time	4	5	3	1	2
Weather	3	2	1	2	3
Avg. Rank	4.5	4.5	3	2	3.5

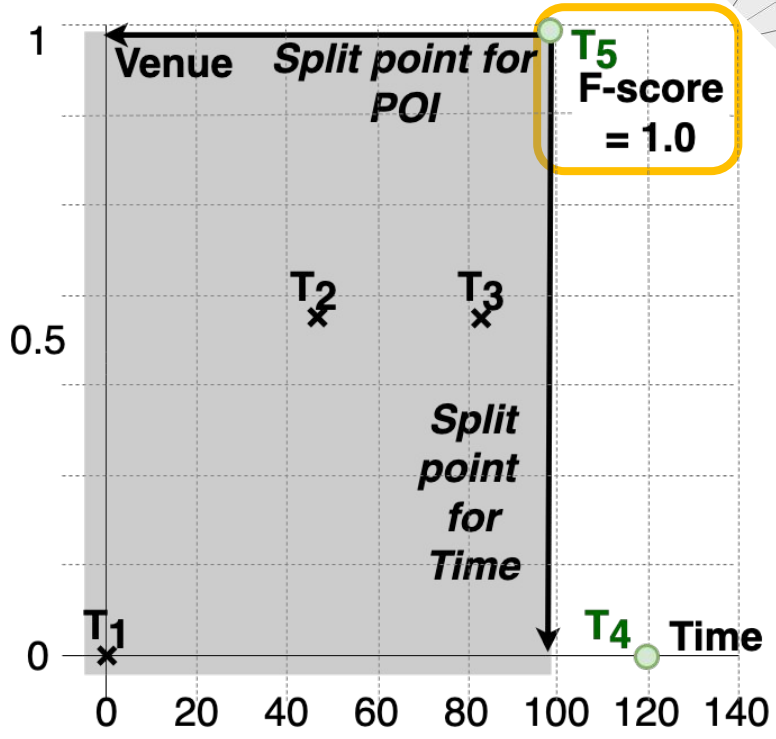


- **Best Alignment**

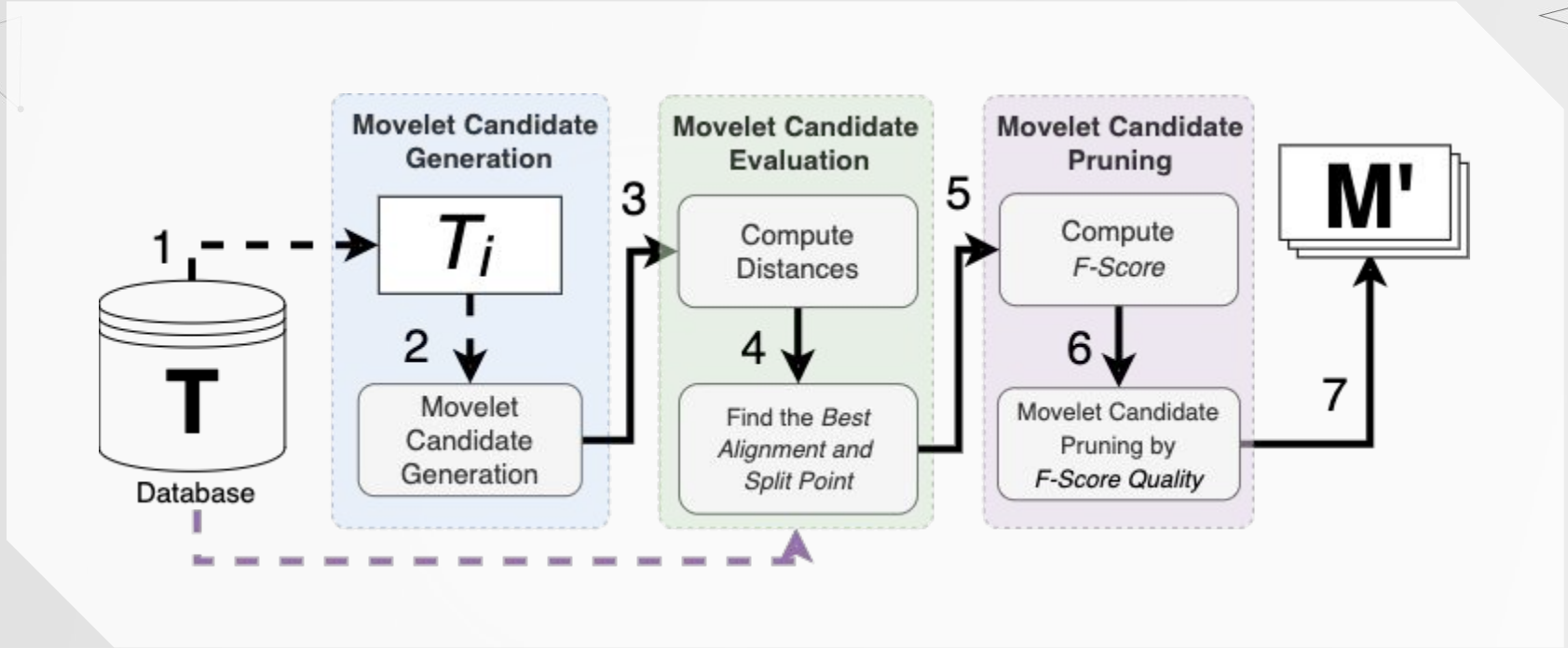
MOVELETS EXTRACTION

Finding the Split Point





MASTERMovelets



(Ferrero et. al, 2020)

MASTERMovelets Complexity

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

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

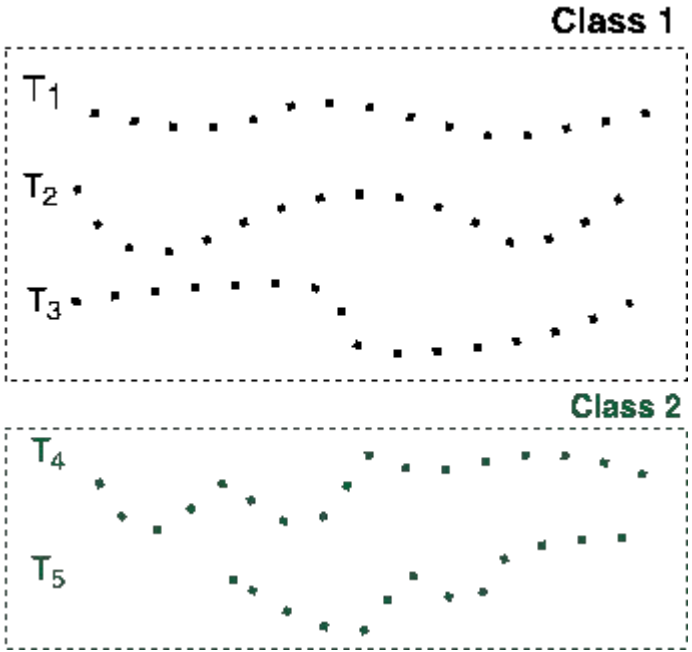
$O(n \times m^2 \times l) \rightarrow$ *Matrix of Distances* *

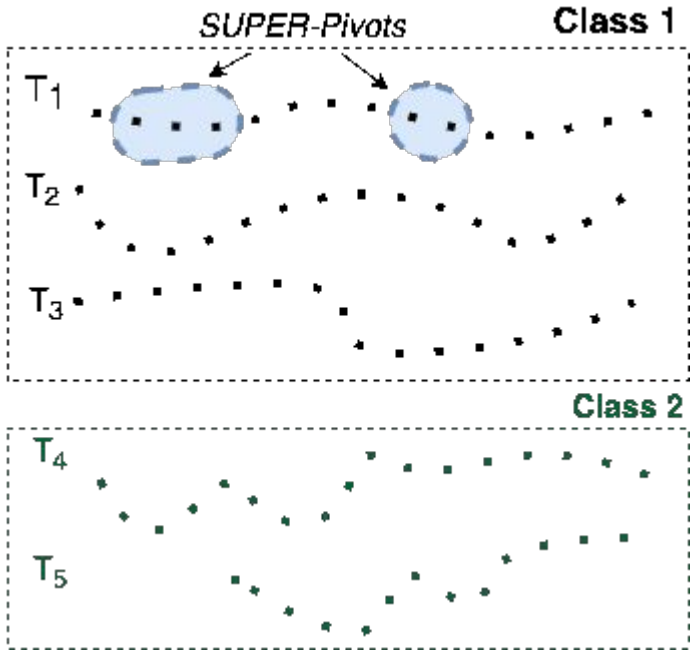
Running Time: the overall time complexity is

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

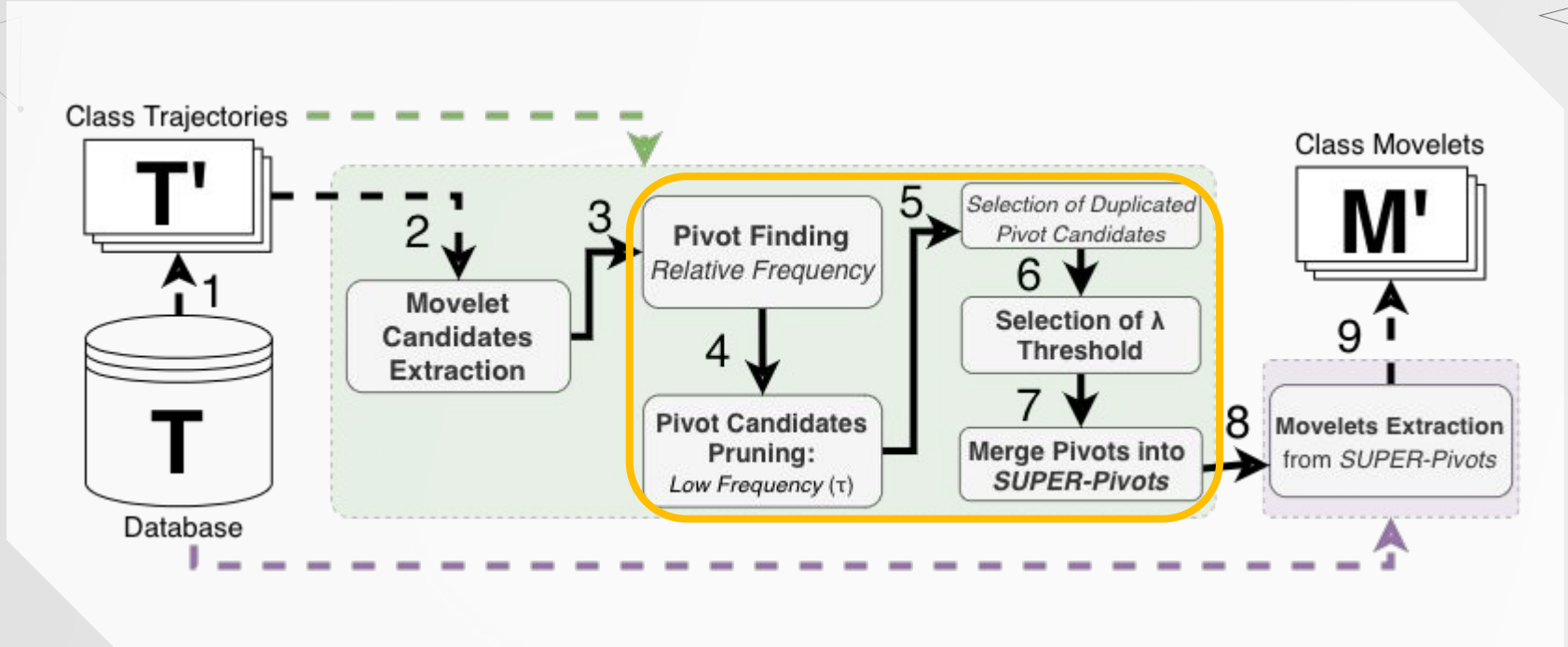
- **Limitations:**
  - Unfeasible for Big Data and high dimensional datasets.

MASTERMovelets x SUPERMovelets





SUPERMovelets



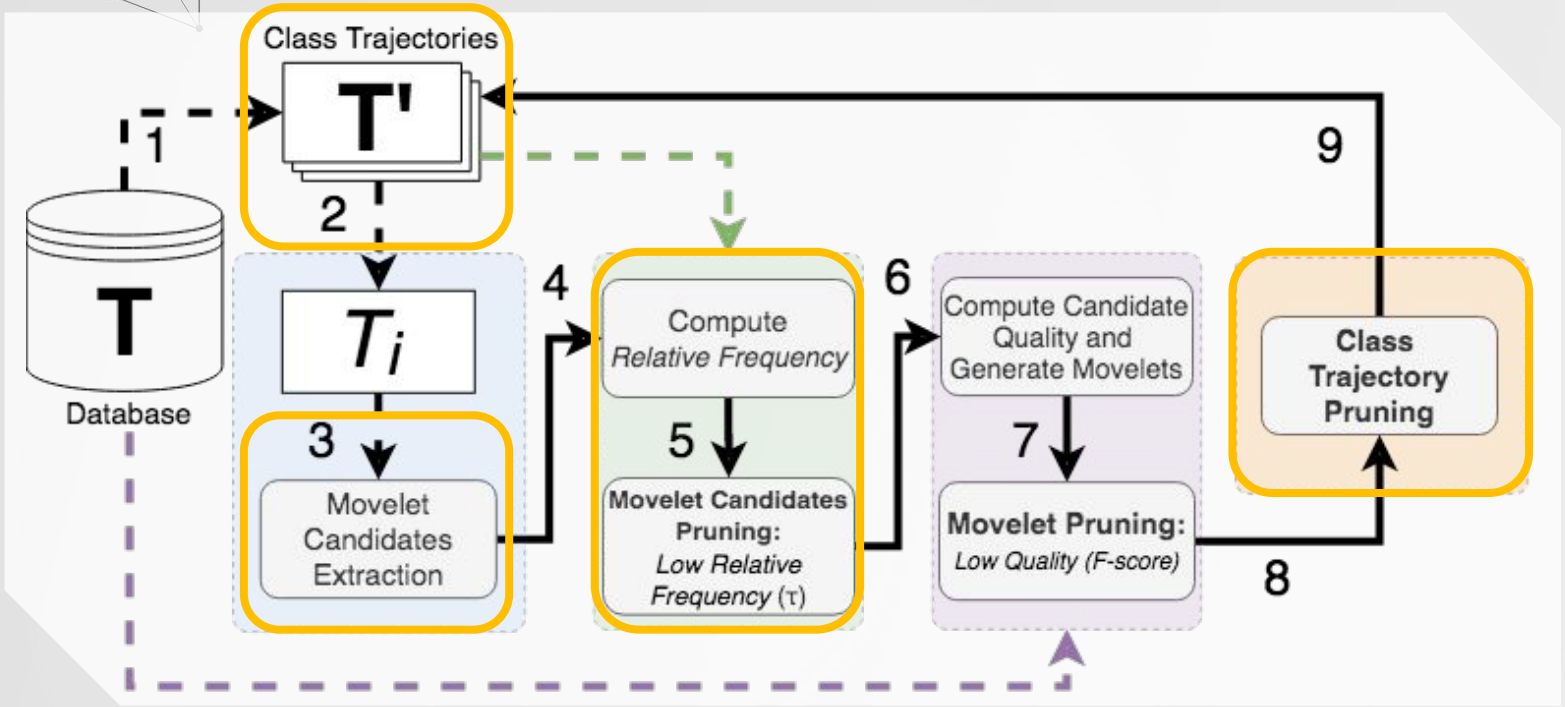
Fast Movelet Extraction and Dimensionality Reduction for Robust Multiple Aspect Trajectory Classification. In Brazilian Conference on Intelligent Systems (BRACIS), 2021.

EXPERIMENTAL RESULTS

Highlights

- **SUPERMovelets movelet extraction:**
  - Movelet extraction at least 50% 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.
- **Limitations:**
  - Unfeasible for Big Data and high dimensional datasets*

HiPerMovelets



HiPerMovelets: high-performance movelet extraction for trajectory classification. International Journal of Geographical Information Science, 2022. DOI: 10.1080/13658816.2021.2018593

EXPERIMENTAL RESULTS

Highlights

- **HiPerMovelets in Multiple Aspect Trajectories Datasets (check-ins):**
  - Run time is up to 10x faster than MASTERMovelets;
  - Higher or same accuracy than MASTERMovelets;;
  - Generates less movelet candidates and movelets.

- **Limitations:**
  - Unfeasible for Big Data and high dimensional datasets.

Open Issues:

1. For which domains Movelets are best suited? (Multiple Aspect Trajectories and Multivariate Time Series)
2. Which are the best strategies to extract movelets? (Optimizing search for Best alignment and Split points)
3. How to improve classification methods? (Employing movelets in Multivariate Time Series Classification, and vice-versa)

REFERENCES

- DABIRI, S.; HEASLIP, K. Inferring transportation modes from GPS trajectories using a convolutional neural network. Transportation Research Part C: Emerging Technologies, Elsevier, v. 86, n. August 2017, p. 360–371, 2018. ISSN 0968090X.
- 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.Discovering Relevant Subtrajectories for Multidimensional Trajectory Classification. 118 p. Tese (PhD thesis) — Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil, 2020.
- 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.
- JI, C. et al. A fast shapelet selection algorithm for time series classification. Computer Networks, Elsevier B.V., v. 148, p. 231–240, 2019.
- LEITE DA SILVA, C. Pivot-based approaches for Movelets and MASTERMovelets Optimizations. 88 p. Master thesis — Universidade Federal de Santa Catarina, Florianópolis,SC, Brasil, 2020.
- 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.
- PORTELA T.T.; CARVALHO J.T.; BOGORNY V. HiPerMovelets: high-performance movelet extraction for trajectory classification, International Journal of Geographical Information Science, 2022. DOI: 10.1080/13658816.2021.2018593.
- PORTELA T.T.; LEITE DA SILVA, C.; CARVALHO J.T.; BOGORNY V. Fast Movelet Extraction and Dimensionality Reduction for Robust Multiple Aspect Trajectory Classification. In Brazilian Conference on Intelligent Systems (BRACIS), 2021.
- 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.

THANKS

Tarlis Tortelli Portela
tarlis.tortelliportela@isti.cnr.it
tarlis@tarlis.com.br

UNIVERSITÀ DI PISA

UNIVERSITÀ DI PISA
CONSIGLIO NAZIONALE DELLE RICERCHE
UNIVERSIDADE FEDERAL DE SANTA CATARINA

Mauriana Pesaresi
Seminar Series 2021/2022

Methods for Movelets Extraction
Multidimensional Sequence Classification

Tarlis Tortelli Portela (tarlis.tortelliportela@isti.cnr.it, tarlis@tarlis.com.br)

Advisors: Anna Bernasconi, Ph.D.
Chiara Renso, Ph.D.
Vania Bogorny, Ph.D.

DATASETS

Multiple-aspect Datasets
Brightkite, Gowalla, Foursquare NYC

- **SPECIFIC**
  high frequency and discriminant behaviors
- **GENERIC**
  high frequency, NOT discriminant behaviors
- **SPATIAL**
  less frequency and discriminant behaviors

Raw Datasets
Animals, GoTrack, Vehicles

- **SPATIO-TEMPORAL**
  Higher GPS granularity

Semantic Datasets

Multivariate Time Series Datasets

Time Series Optimization

EARLY ABANDONING

(YE; KEOGH, 2011)
(MUEEN; KEOGH; YOUNG, 2011)

SAX

(RAKTHANMANON; KEOGH, 2013)

SAMPLING THE DATASET

(JI et al., 2019).

ANALYSIS OF VARIANCE A(NOVA)

(ZUO; ZEITOUNI; TAHER, 2018)

LOCAL FISHER DISCRIMINANT ANALYSIS (LFDA)

(ZHANG et al., 2018)

PROPOSED OPTIMIZATIONS

A) Pruning based on repetition

(prunes candidates that doesn't repeat in the trajectory)

B) Pruning based on frequency

(prunes less frequent candidates)

C) Random selection * for baseline

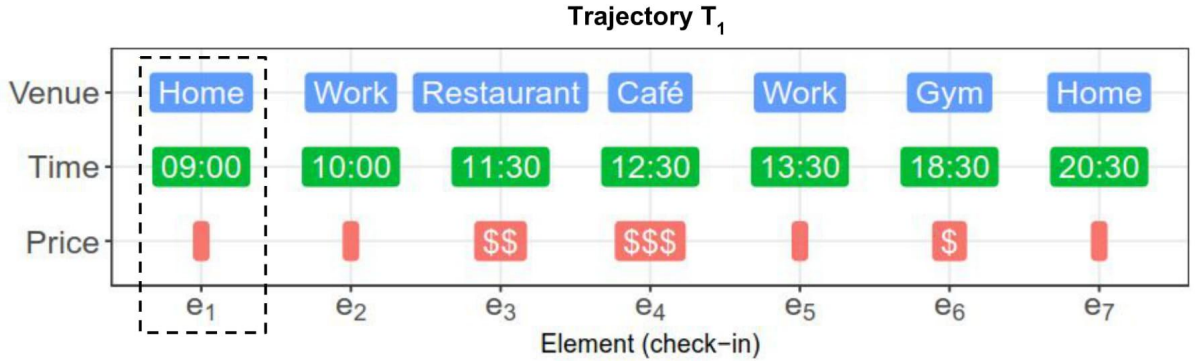
(randomly select candidates to evaluate as movelets)

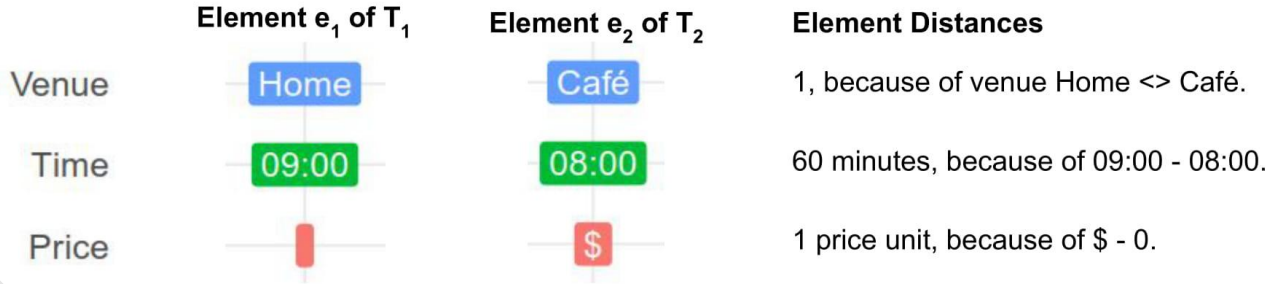
OBJECTIVE

Propose new methods to optimize the *movelets discovery for trajectory classification*.

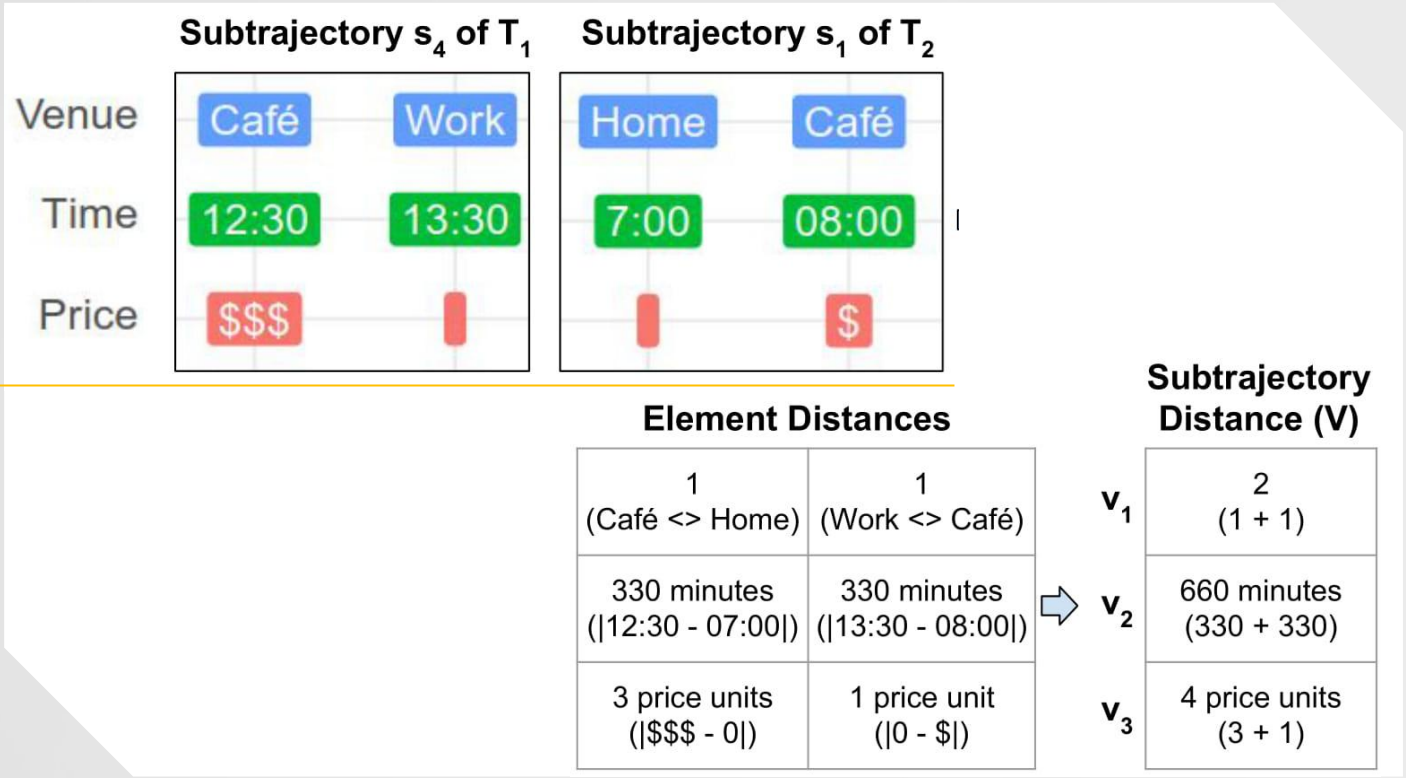
- Propose techniques for trajectory classification based on reducing the search space, dimensionality and number of comparisons;
  - An algorithm to reduce the search space for extracting movelets;
  - An algorithm to reduce the number of dimensions for extracting movelets;
  - An algorithm that uses a multidimensional index to reduce the number of comparisons for extracting movelets;
- Experiments for validating the proposed method and MASTERMovelets (scalability, comparisons of accuracy, processing time, classification times and number of movelets).

Basic Concepts:
Element Distances

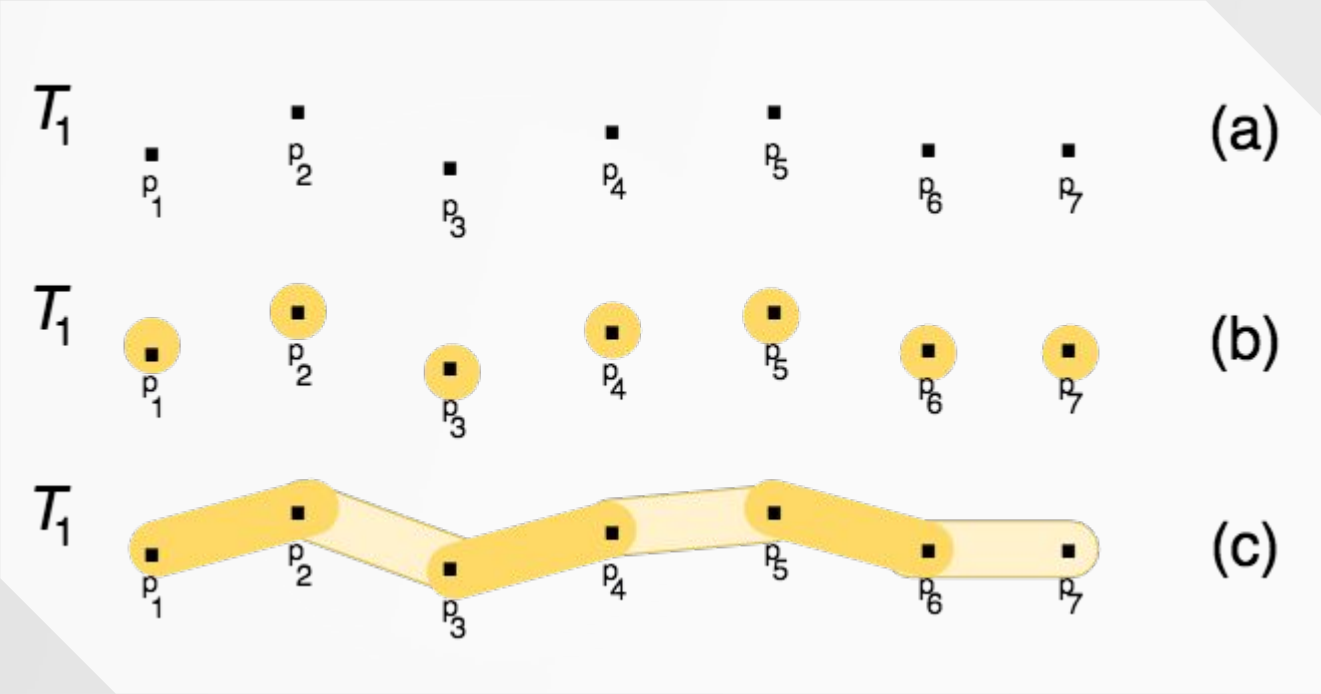




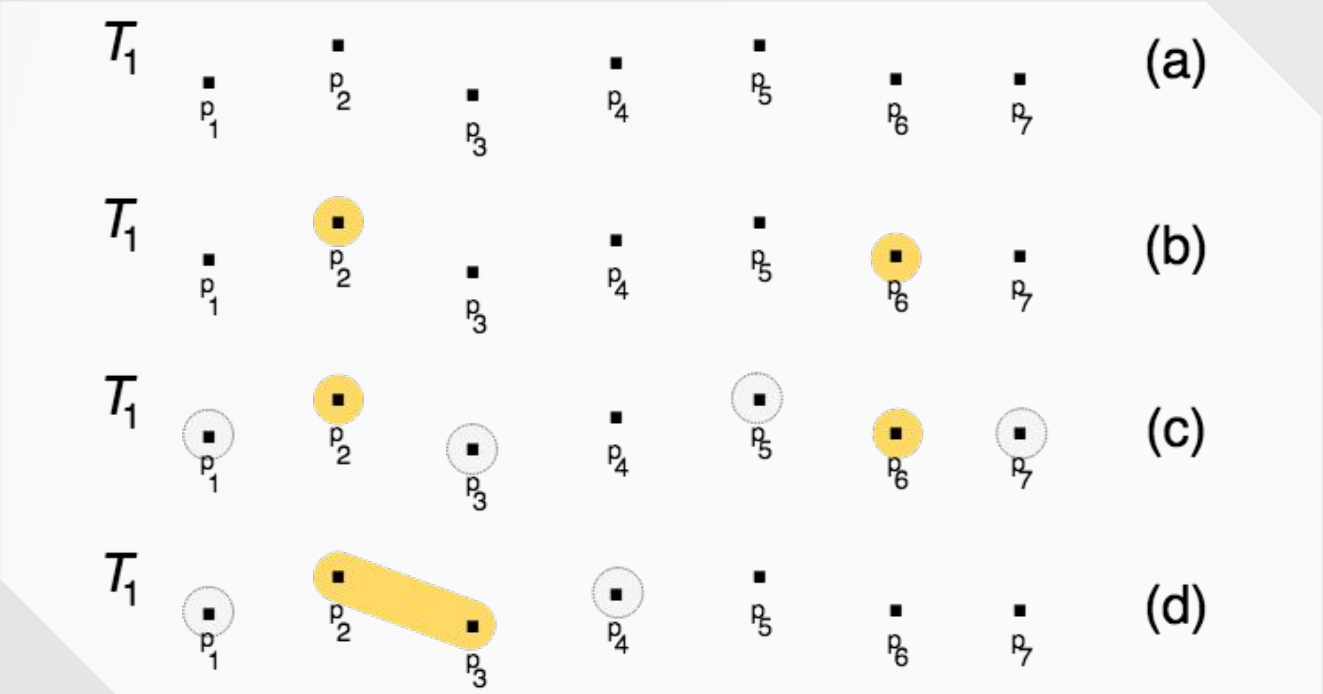
Basic Concepts: *Subtrajectory Distances*



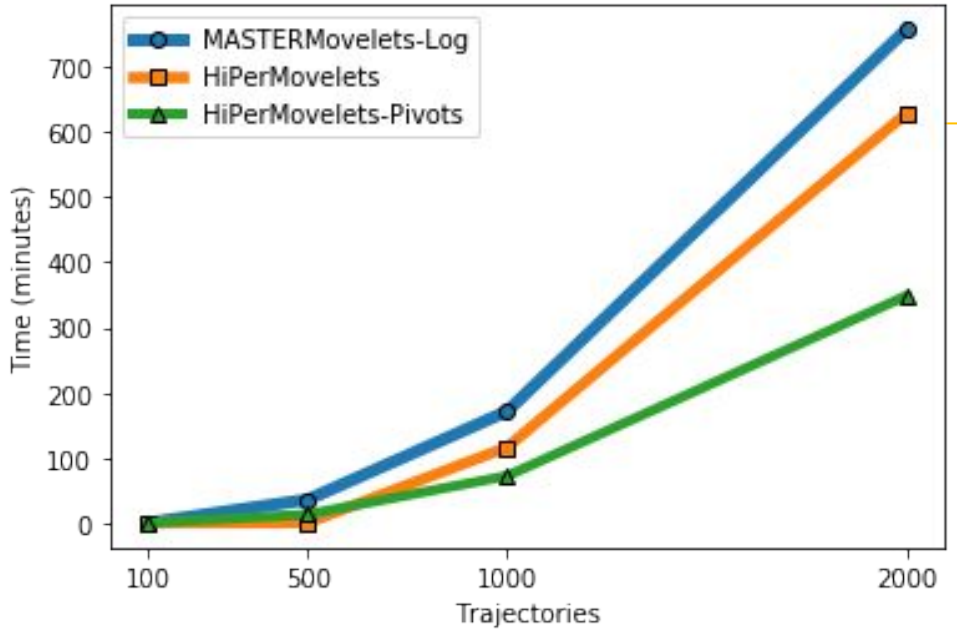
HiPerMovelets



HiPerMovelets
Pivots



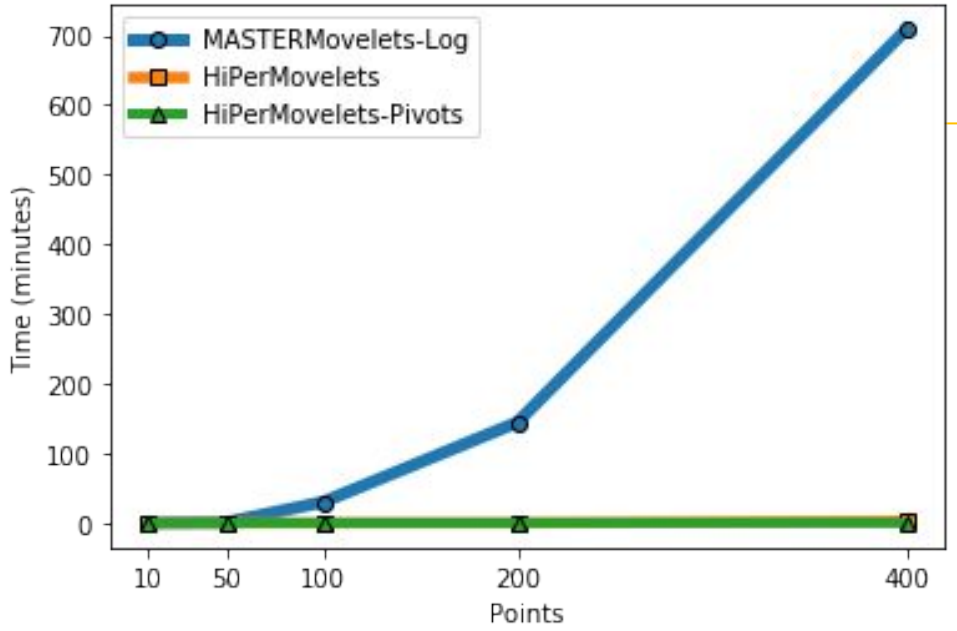
EXPERIMENTAL RESULTS



Scalability:
Number of Trajectories

All experiments: faster as trajectories are added

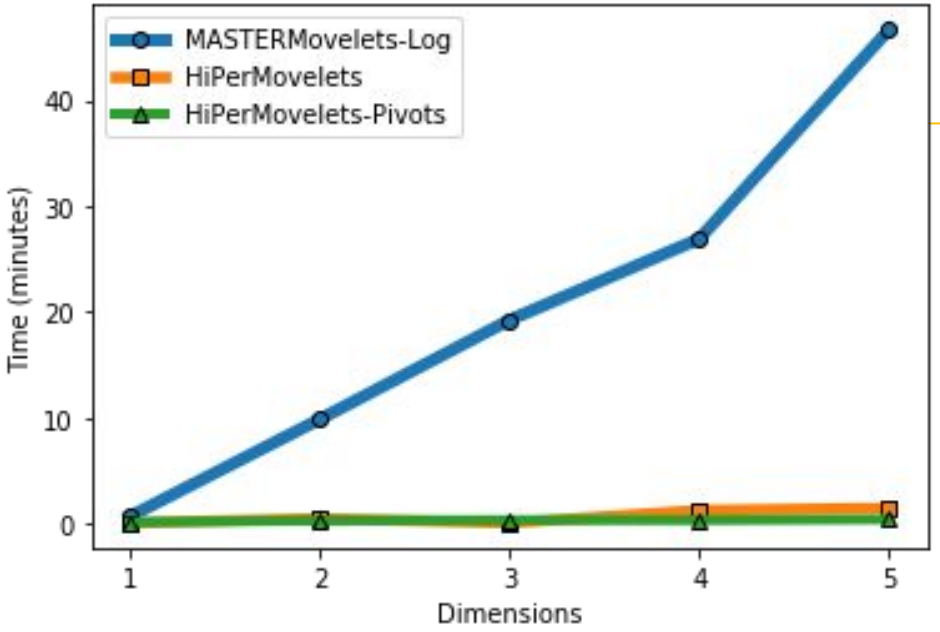
EXPERIMENTAL RESULTS



Scalability:
Number of Points

All experiments:
faster as points are added

EXPERIMENTAL RESULTS



Scalability: Number of Dimensions

All experiments: faster as dimensions are added