Exploring Frequency-based Approaches for Efficient Trajectory Classification

Francisco Vicenzi
Lucas May Petry
Camila Leite da Silva
Luis Otavio Alvares
Vania Bogorny (vania@inf.ufsc.br)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)
Universidade Federal de Santa Catarina (UFSC),
Florianópolis, Brazil

35th ACM Symposium on Applied Computing

Outline

- Introduction and Motivation
- Problem Definition
- Objective and Contributions
- Proposal
- Experimental Evaluation
- Conclusion and Future Work

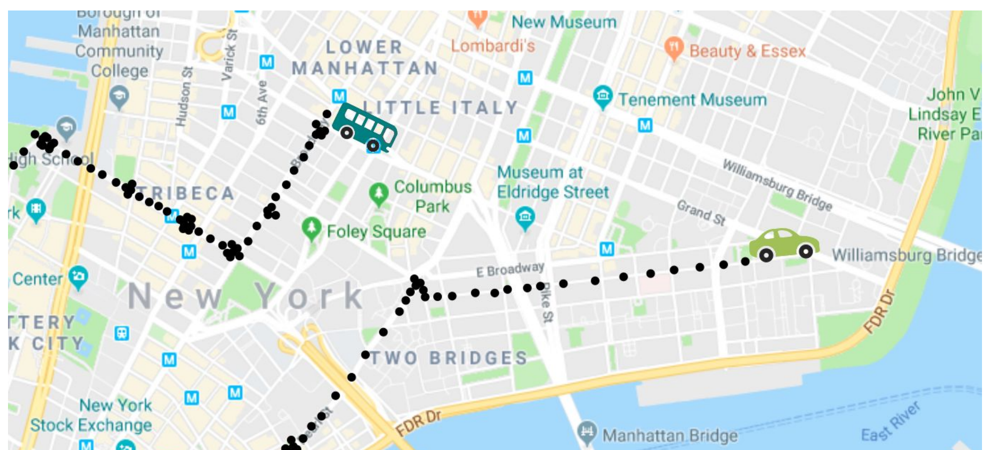
UNIVERSIDADE FEDERAL
DE SANTA CATARINA

Introduction

- ➢ Era of movement tracking, and with the Corona virus, tracking people became important, urgent, and easy as never before in human history

For many years, moving object traces were called Raw Trajectories

- Sequence of points located in space and time
- Data are dense and have so semantics

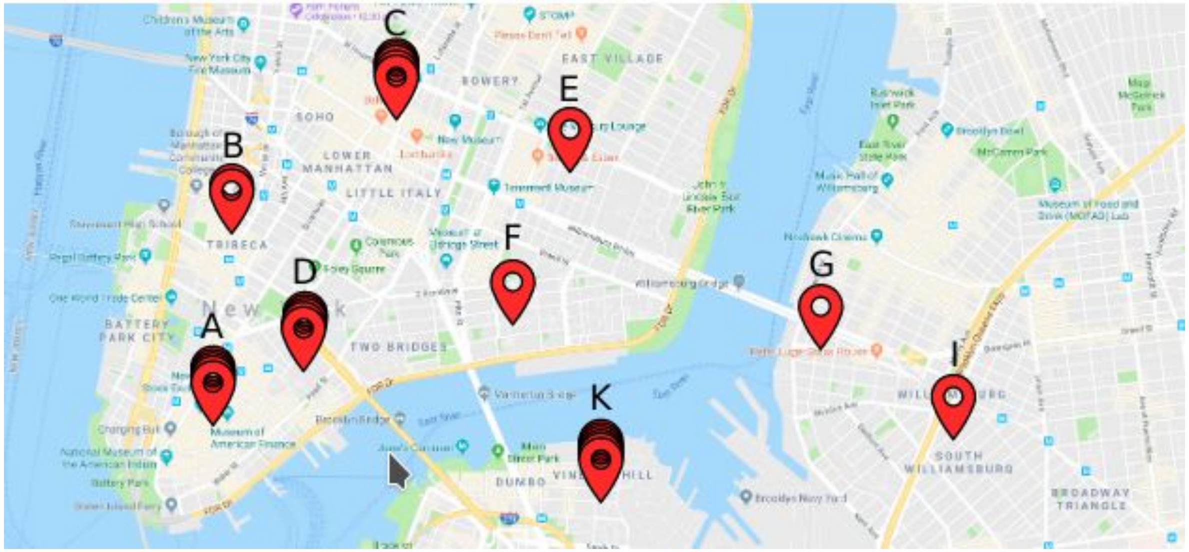


UNIVERSIDADE FEDERAL
DE SANTA CATARINA

Introduction

After 2007 emerged the concept of *Semantic Trajectory*

- Sequence of points located in space and time enriched with **semantic** information
- The figure shows examples of social media trajectories, where each point is a visited place (called Point of Interest (POI)), and POI related attributes as *place category*, *price* and *rating*,



Introduction

An important task in trajectory data mining is *Classification*

It consists of categorizing a moving object based on its trajectories.

Applications

- Classifying animal species
- Transportation mode inference (car, bus, taxi, etc);
- Profiling human behavior
- Individuals that are potential virus spreaders

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Problem Definition

Given a large dataset of trajectories, how can we ***efficiently*** classify the moving objects based on their trajectories?

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Related Work

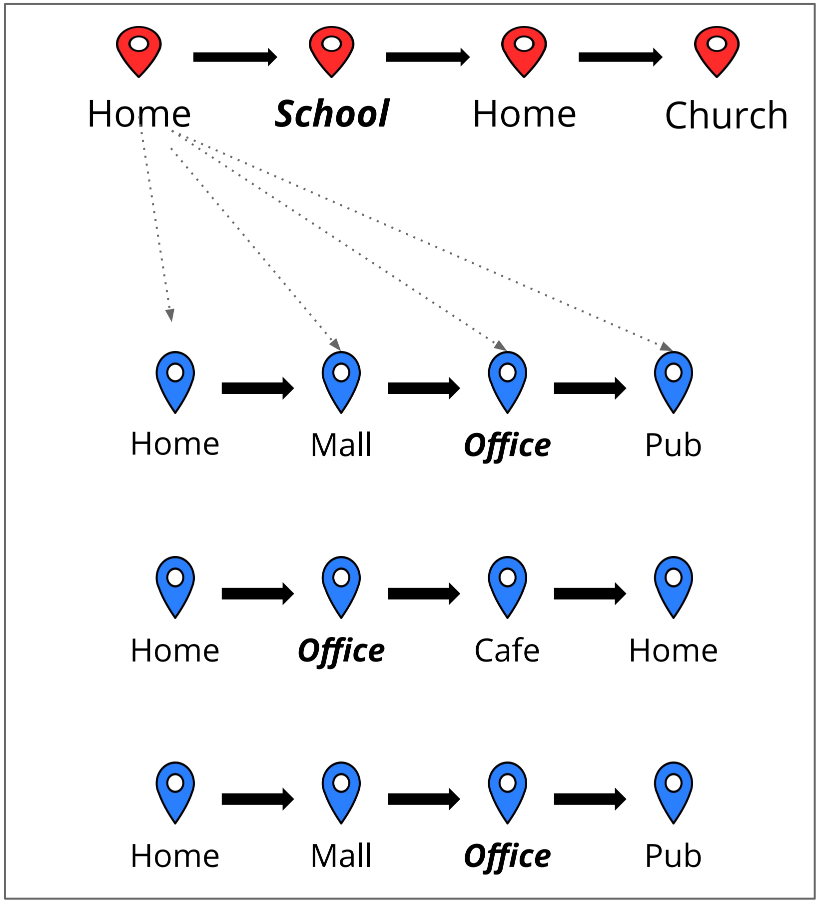
Work	Space	Time	Semantics	Frequency -based	Trajectory features used as input for the classifier
TraClass (Lee et al. 2008)	✓			✓	Subtrajectories
Patel et al. (2012)	✓	✓		✓	Subtrajectories
Dodge et al. (2009)	✓	✓			Features (speed, acceleration, ...)
Soleymani et al. (2014)	✓	✓			Space partitions + Temporal descriptive statistics
Etemad et al. (2018)					Features (speed, acceleration…)
Lee et al. (2011)			✓		Frequent segments (street names)
Bi-TULER (Gao et al. 2017)			✓		POI embeddings
TULVAE (Zhou et al. 2018)			✓		POI embeddings
Movelets (Ferrero et al. 2018)	✓	✓	✓	✓	Subtrajectories

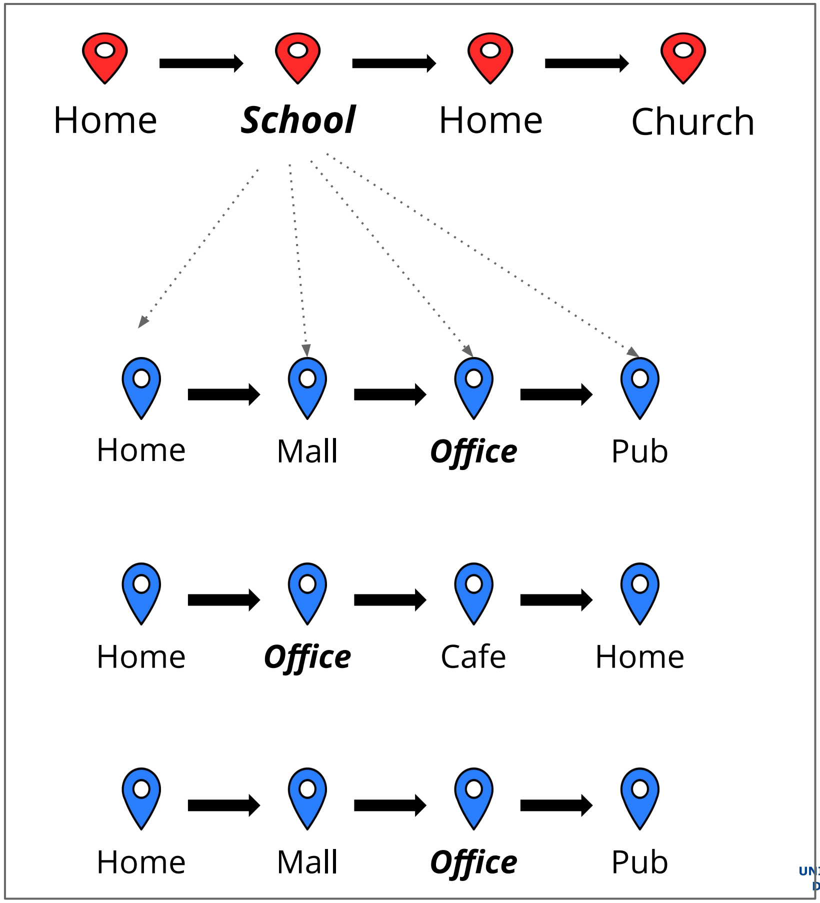
- **All methods extract features from trajectories and use traditional classifiers**
- **Some are limited to a specific application, as transportation mode inference**
- **Movelets is generic for any application and is the best method so far in terms of accuracy**

UNIVERSIDADE FEDERAL
DE SANTA CATARINA

Problem definition

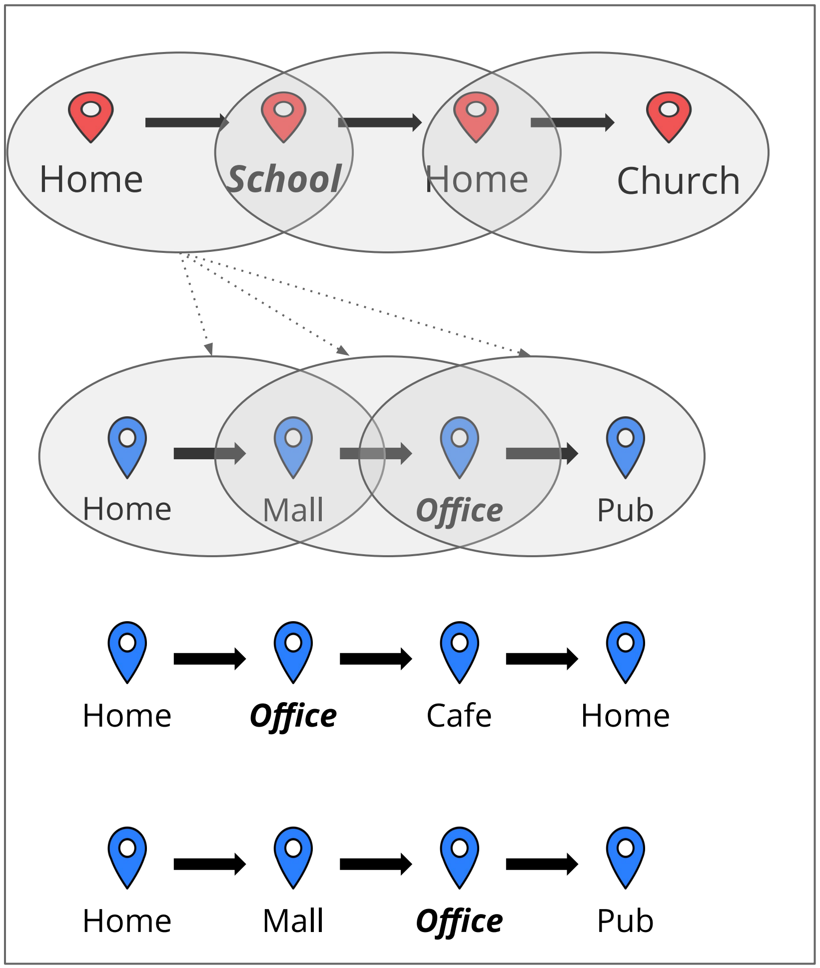
Movelets extracts every possible subtrajectory and computes the distance to all trajectories in the dataset to find the most discriminant subtrajectories

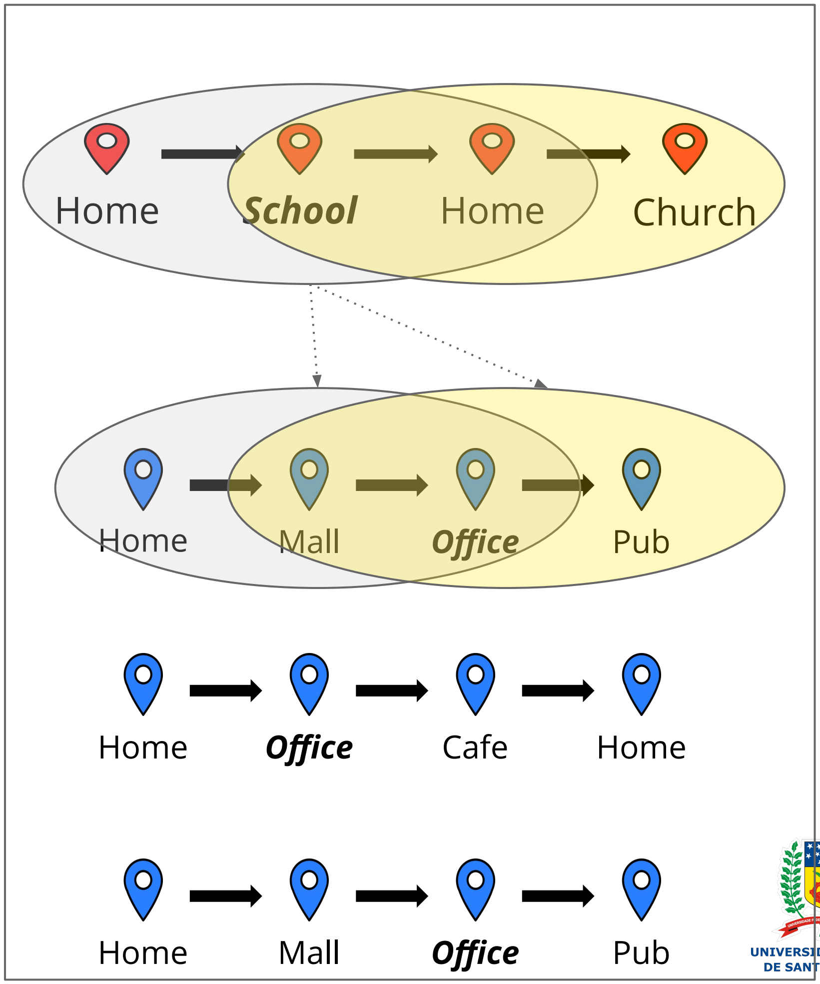




Problem definition

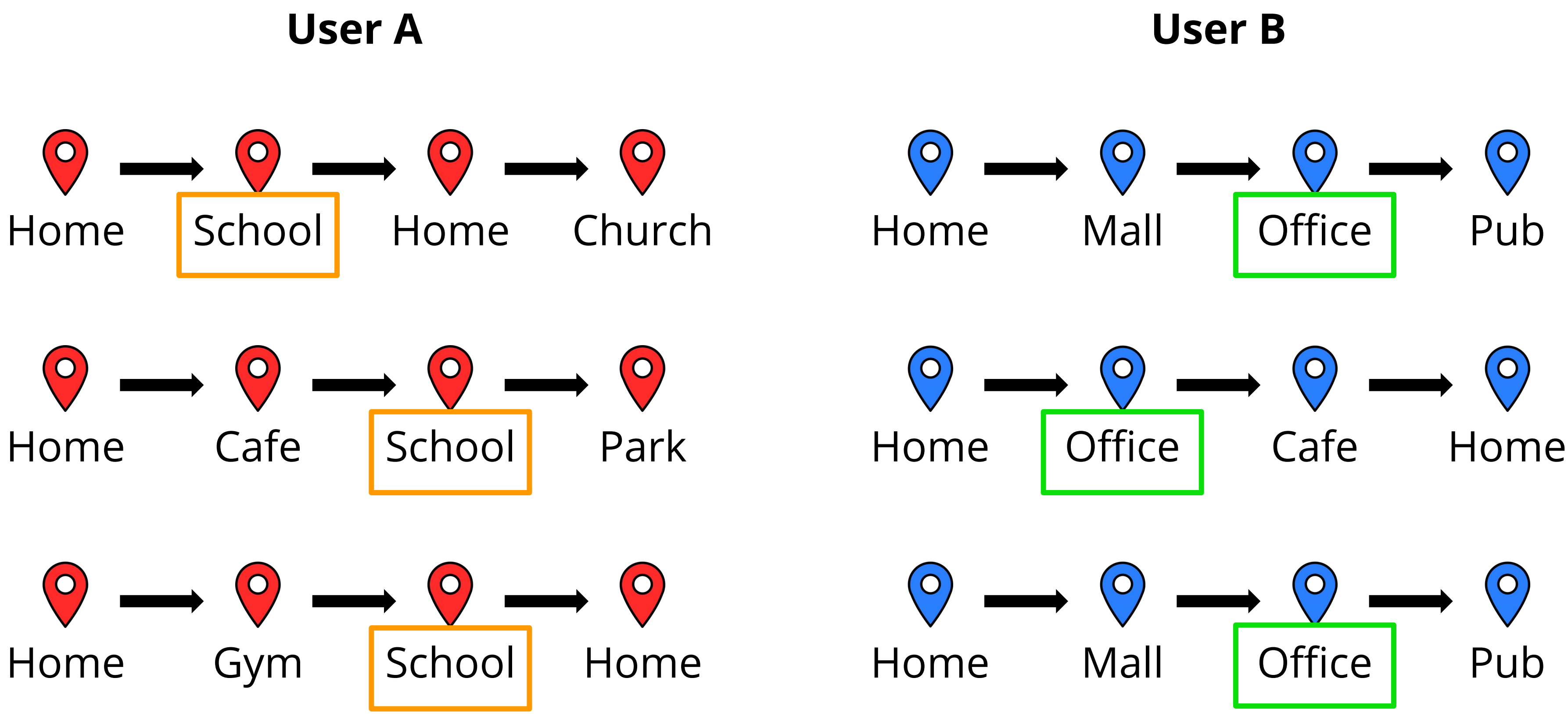
The approach is not scalable for big data





UNIVERSIDADE FEDERAL DE SANTA CATARINA

Problem Definition



Users show a certain frequency of some visited POIs

Objective

Propose frequency-based methods for semantic trajectory classification that are faster and achieve similar or better results than the state of the art

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Contributions

Three complementary fast approaches based on the **frequency** of **visits**:

- **POI-F** (POI Frequency)
- **NPOI-F** (Normalized POI Frequency)
- **WNPOI-F** (Weighted Normalized POI-Frequency)

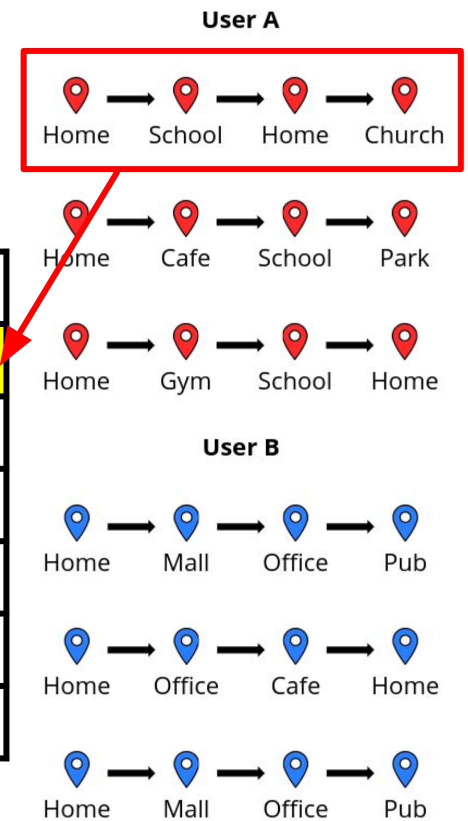
UNIVERSIDADE FEDERAL DE SANTA CATARINA

Proposal

POI-Frequency (POI-F): Given a trajectory *T* and a POI *P*, the frequency of *P* in *T* is computed as

$$\text{POI-F}(P, T) = \text{the number of visits of P in T.}$$

	Home	School	Church	Cafe	Park	Gym	Mall	Office	Pub	Class
A1	**2**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**A**
A2	1	1	0	1	1	0	0	0	0	**A**
A3	2	1	0	0	0	1	0	0	0	**A**
B1	1	0	0	0	0	0	1	1	1	**B**
B2	2	0	0	1	0	0	0	1	0	**B**
B3	1	0	0	0	0	0	1	1	1	**B**



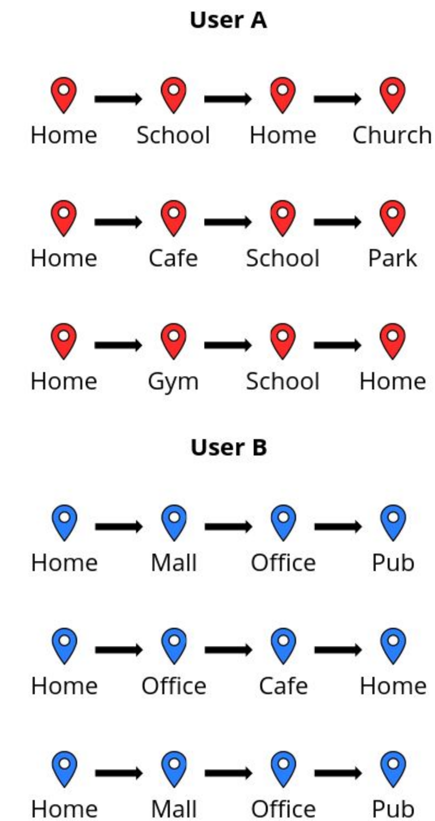
The trajectories become a frequency matrix, where each row is a trajectory and the columns are all POIs in the database

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Proposal

POI-Frequency (POI-F): Given a trajectory *T* and a POI *P*, the frequency of *P* in *T* is computed as

$$\text{POI-F}(P, T) = \text{the number of visits of P in T.}$$



	Home	School	Church	Cafe	Park	Gym	Mall	Office	Pub	Class
A1	**2**	**1**	**1**	0	0	0	0	0	0	**A**
A2	**1**	**1**	0	**1**	**1**	0	0	0	0	**A**
A3	**2**	**1**	0	0	0	**1**	0	0	0	**A**
B1	**1**	0	0	0	0	0	**1**	**1**	**1**	**B**
B2	**2**	0	0	**1**	0	0	0	**1**	0	**B**
B3	**1**	0	0	0	0	0	**1**	**1**	**1**	**B**

The red cells represent the POIs visited by user A and the blue ones represent user B

UNIVERSIDADE FEDERAL DE SANTA CATARINA

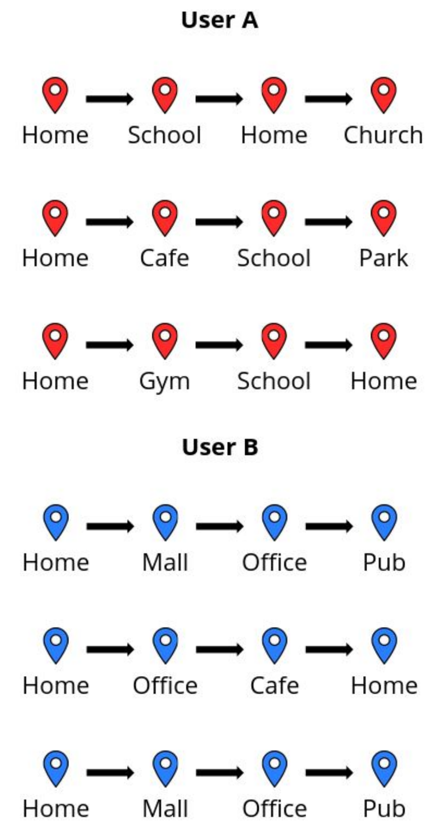
Proposal

POI-Frequency (POI-F): Given a trajectory *T* and a POI *P*, the frequency of *P* in *T* is computed as

$$\text{POI-F}(P, T) = \text{the number of visits of P in T.}$$

	Home	School	Church	Cafe	Park	Gym	Mall	Office	Pub	Class
A1	2	**1**	1	0	0	0	0	0	0	**A**
A2	1	**1**	0	1	1	0	0	0	0	**A**
A3	2	**1**	0	0	0	1	0	0	0	**A**
B1	1	0	0	0	0	0	**1**	**1**	**1**	**B**
B2	2	0	0	1	0	0	0	**1**	0	**B**
B3	1	0				0	**1**	**1**	**1**	**B**

Discriminative patterns



UNIVERSIDADE FEDERAL DE SANTA CATARINA

Proposal

Normalized POI Frequency (NPOI-F). Given by the proportion of the frequency of a POI (P) in a trajectory (T) and the size of the trajectory

$$\text{NPOI-F}(P,T) = \frac{\text{POI-F(P,T)}}{\text{\# of points in T}}$$

- Useful for comparing the behavior of users in trajectories of different lengths.

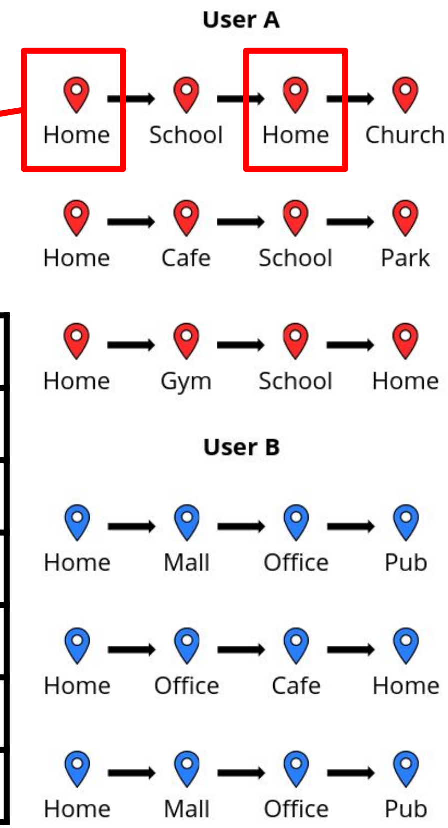
UNIVERSIDADE FEDERAL DE SANTA CATARINA

Proposal

Our previous example computing the **NPOI-F**:

$$\text{NPOI-F}(P,T) = \frac{\text{POI-F(P,T)}}{\#\text{ of points in T}}$$

$$\text{NPOI-F}(Home, A1) = \frac{2}{4} = 0.5$$



	Home	School	Church	Cafe	Park	Gym	Mall	Office	Pub	Class
A1	$\boxed{0.50}$	0.25	0.25	0	0	0	0	0	0	A
A2	0.25	0.25	0	0.25	0.25	0	0	0	0	A
A3	0.50	0.25	0	0	0	0.25	0	0	0	A
B1	0.25	0	0	0	0	0	0.25	0.25	0.25	B
B2	0.50	0	0	0.25	0	0	0	0.25	0	B
B3	0.25	0	0	0	0	0	0.25	0.25	0.25	B

Visits now are given as percentages according to the whole trajectory (each line sums up to 1)

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Proposal

Weighted Normalized POI Frequency. Considers the frequency wrt the number of users in the database and the number of users that have visited a place.

$$\text{WNPOI-F}(P, T, D) = \text{NPOI-F} * log\left(\frac{\text{\# of users in D}}{\text{\# of users that have visited P}}\right)$$

Useful for penalizing visits that many users (classes) have in common.

In the previous example, the frequencies of the POI Home would become zero because they do not discriminate the users.

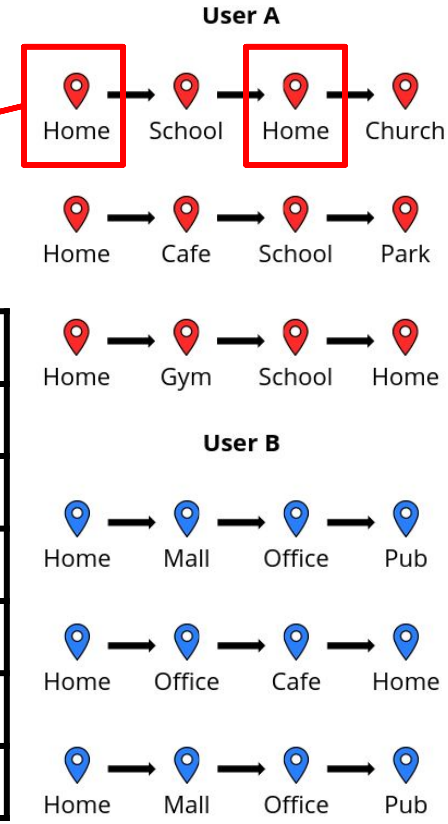
UNIVERSIDADE FEDERAL DE SANTA CATARINA

Proposal

Our previous example computed by using **WNPOI-F**:

$$\text{WNPOI-F}(P,T,D) = \text{NPOI-F} * log\left(\frac{\text{\# of users in D}}{\text{\# of users that have visited P}}\right)$$

$$\text{WNPOI-F}(Home, A1, 6) = \frac{2}{4} * log\left(\frac{2}{2}\right) = 0.5 * 0 = 0$$



	Home	School	Church	Cafe	Park	Gym	Mall	Office	Pub	Class
A1	0	0.07	0.07	0	0	0	0	0	0	A
A2	0	0.07	0	0	0.10	0	0	0	0	A
A3	0	0.07	0	0	0	0.07	0	0	0	A
B1	0	0	0	0	0	0	0.07	0.07	0.07	B
B2	0	0	0	0	0	0	0	0.07	0	B
B3	0	0	0	0	0	0	0.07	0.07	0.07	B

They are penalized because they do not discriminate the classes

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Experimental Evaluation

- ➢ Evaluation with three real **datasets** from three location-based social networks: Brightkite, Gowalla, and Foursquare;
- ➢ **Comparison** with Bi-TULER, TULVAE, and Movelets;
- ➢ **Metrics**: ACC@1, ACC@5 and Macro-F1;
- ➢ Classification algorithm: MultiLayer Perceptron (MLP) - determined by preliminary experiments;
  - Hidden layer with 100 units;
  - Trained with Adam optimizer and a learning rate of 10^{-3};
  - batch size of 64;
  - dropout rate of 0.5.

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Datasets

Datasets used by previous methods and that have lat, lon, time and semantic dimensions

Dataset	Description	
Brightkite	Min - Max Traj. Size	10 - 50
	Number of Trajectories	7911
	Number of Points	130494
	Attributes	Lat, Lon, POI, Time, Weekday, Class
	Class	Users (300)
Gowalla	Min - Max Traj. Size	10 - 50
	Number of Trajectories	5329
	Number of Points	98158
	Attributes	Lat, Lon, POI, Time, Weekday, Class
	Class	Users (300)
Foursquare NYC	Min - Max Traj. Size	10 - 144
	Number of Trajectories	3079
	Number of Points	66962
	Attributes	Lat, Lon, POI, Time, Weekday, Weather, Price, Rating, Class
	Class	Users (193)

UNIVERSIDADE FEDERAL
DE SANTA CATARINA

Results with the three datasets (Accuracy)

Brightkite

Method	ACC@1	ACC@5	Macro-F1
Bi-TULER [Gao et al., 2017]	90.64	95.55	87.92
MOVELETS [Ferrero et al., 2018]	90.55	93.79	89.41
TULVAE [Zhou et al., 2018]	88.41	92.15	83.63
POI-F	<u>95.29</u>	**97.98**	<u>93.80</u>
NPOI-F	**95.34**	**97.98**	**93.91**
WNPOI-F	95.13	<u>97.94</u>	93.68

Gowalla

Method	ACC@1	ACC@5	Macro-F1
Bi-TULER [Gao et al., 2017]	66.15	78.36	63.26
MOVELETS [Ferrero et al., 2018]	91.44	94.04	90.25
TULVAE [Zhou et al., 2018]	67.94	78.76	64.91
POI-F	<u>93.22</u>	<u>97.62</u>	**92.54**
NPOI-F	**93.46**	**97.68**	<u>91.91</u>
WNPOI-F	93.11	97.45	91.66

Foursquare

Method	ACC@1	ACC@5	Macro-F1
Bi-TULER [Gao et al., 2017]	48.20	67.38	40.56
MOVELETS [Ferrero et al., 2018]	97.66	98.93	96.45
TULVAE [Zhou et al., 2018]	54.33	73.81	46.54
POI-F	<u>98.05</u>	**99.41**	97.29
NPOI-F	**98.24**	**99.41**	<u>97.37</u>
WNPOI-F	**98.24**	**99.41**	**97.54**

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Results with Generalized Foursquare (Accuracy)

Making the classification problem more difficult - Generalizing the **POI name** to **POI category (e.g. Ibis Hotel → Hotel)**

Method	ACC@1	ACC@5	Macro-F1
Bi-TULER [Gao et al., 2017]	33.50	60.76	28.29
MOVELETS [Ferrero et al., 2018]	35.34	65.43	29.39
TULVAE [Zhou et al., 2018]	32.81	61.15	28.14
POI-F	**36.22**	**67.96**	**31.12**
NPOI-F	35.24	67.47	28.30
WNPOI-F	34.07	67.47	30.26

Significant accuracy reduction of ACC@1

36,22 % best

Generalized Foursquare considering **additional attributes**

Method	Attributes	ACC@1
MOVELETS [Ferrero et al., 2018]	POI Category + Time	44.59
POI-F	POI Category + Rating	71.56
NPOI-F	POI Category + Price + Rating	**74.09**
WNPOI-F	POI Category + Price + Rating	**74.09**

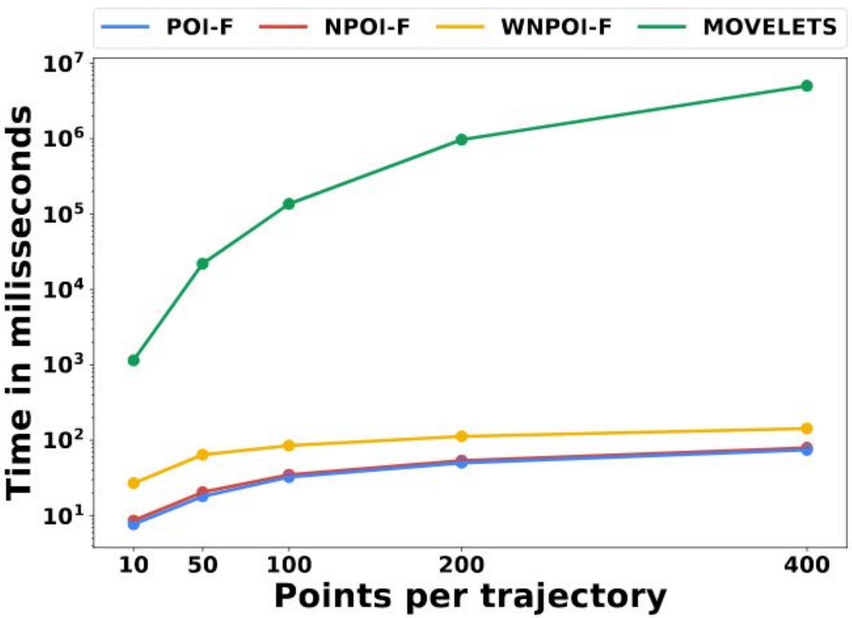
Accuracy increased to 74,09%

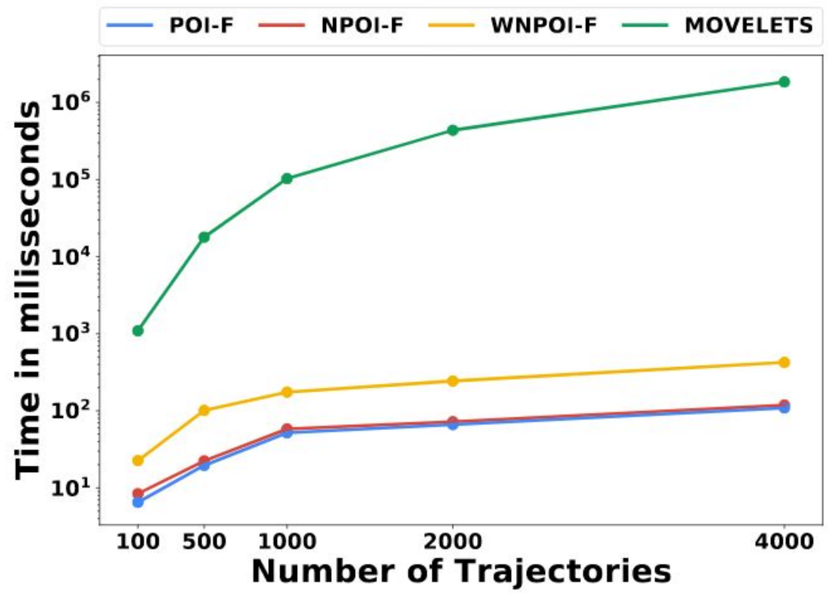
UNIVERSIDADE FEDERAL DE SANTA CATARINA

Scalability Analysis

Increasing the number of trajectory points

Increasing the number of trajectories





UNIVERSIDADE FEDERAL
DE SANTA CATARINA

Conclusion and Future Work

- ➢ Simple, effective and efficient method for **semantic trajectory classification**;
- ➢ Much higher **accuracy** than Bi-TULER and TULVAE, and equal or higher than MOVELETS;
- ➢ Approach feasible to be used for **big semantic trajectory data** due to its efficient processing time;
- ➢ **Future works**: consider the sequence of POIs.

UNIVERSIDADE FEDERAL DE SANTA CATARINA

Acknowledgements

Brazilian agencies

CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Finance Code 001),

CNPQ (Conselho Nacional de Desenvolvimento Científico e Tecnológico),

FAPESC (Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina - Project Match - Co-financing of **H2020** Projects - Grant 2018TR 1266).

UNIVERSIDADE FEDERAL
DE SANTA CATARINA

Thank You!

Contact: vania.bogorny@ufsc.br

UNIVERSIDADE FEDERAL
DE SANTA CATARINA

References

1. Marco Aurelio Beber, Carlos Andres Ferrero, Renato Fileto, and Vania Bogorny.2017. **Individual and Group Activity Recognition in Moving Object Trajectories**. JIDM8, 1 (2017), 50–66.
2. Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. **Friendship and mobility: user movement in location-based social networks.** In Proceedings of the 17th ACMSIGKDD international conference on Knowledge discovery and data mining. ACM,1082–1090.
3. Lucas Andre de Alencar, Luis Otávio Alvares, Chiara Renso, Alessandra Raffaetà, and Vania Bogorny. 2015. **A Rule-based Method for Discovering Trajectory Profiles**. In The 27th International Conference on Software Engineering and Knowledge Engineering, SEKE 2015, Wyndham Pittsburgh University Center, Pittsburgh, PA,USA, July 6-8, 2015, Haiping Xu (Ed.). KSI Research Inc. and Knowledge Systems Institute Graduate School, 244–249.
4. Somayeh Dodge, Robert Weibel, and Ehsan Forootan. 2009. **Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects.** Computers, Environment and Urban Systems 33, 6 (2009),419–434.
5. Mohammad Etemad, Amílcar Soares Júnior, and Stan Matwin. 2018. **Predicting Transportation Modes of GPS Trajectories using Feature Engineering and Noise Removal.** In Advances in Artificial Intelligence: 31st Canadian Conference on Artificial Intelligence, Canadian AI 2018, Toronto, ON, Canada, May 8–11, 2018, Proceedings 31. Springer, 259–264.
6. Carlos Andres Ferrero, Luis Otavio Alvares, Willian Zalewski, and Vania Bogorny.2018. **MOVELETS: Exploring Relevant Subtrajectories for Robust Trajectory Classification**. In Proceedings of the 33rd ACM/SIGAPP Symposium on Applied Computing, Pau, France. 849–856.

UNIVERSIDADE FEDERAL
DE SANTA CATARINA

References

7. Qiang Gao, Fan Zhou, Kunpeng Zhang, Goce Trajcevski, Xucheng Luo, and Fengli Zhang. 2017. **Identifying Human Mobility via Trajectory Embeddings**. In Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17). AAAI Press, 1689–1695.
8. Marta C. González, César A. Hidalgo, and Albert-László Barabási. 2008. **Understanding individual human mobility patterns**. Nature 453, 7196 (June 2008),779–782.
9. Jae-Gil Lee, Jiawei Han, Xiaolei Li, and Hector Gonzalez. 2008. **TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering**. Proceedings of the VLDB Endowment 1, 1 (2008), 1081–1094.
10. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. **Efficient estimation of word representations in vector space**. arXiv preprint arXiv:1301.3781(2013).
11. Dhaval Patel, Chang Sheng, Wynne Hsu, and Mong Li Lee. 2012. **Incorporating duration information for trajectory classification**. In Data Engineering (ICDE),2012 IEEE 28th International Conference on. IEEE, 1132–1143.
12. Spiridoula Tragopoulou, Iraklis Varlamis, and Magdalini Eirinaki. 2014. **Classification of movement data concerning user's activity recognition via mobile phones**. In Proceedings of the 4th International Conference on Web Intelligence,Mining and Semantics (WIMS14). ACM, 42: 1–6.
13. Iraklis Varlamis. 2015. **Evolutionary data sampling for user movement classification**. In Congress on Evolutionary Computation (CEC), 2015. IEEE.
14. Xiangye Xiao, Yu Zheng, Qiong Luo, and Xing Xie. 2014. **Inferring social ties between users with human location history**. Journal of Ambient Intelligence and Humanized Computing 5, 1 (01 Feb 2014), 3–19.

UNIVERSIDADE FEDERAL DE SANTA CATARINA

References

15. Zhibin Xiao, Yang Wang, Kun Fu, and Fan Wu. 2017. **Identifying Different Transportation Modes from Trajectory Data Using Tree-Based Ensemble Classifiers**. ISPRS Int. J. Geo-Inf.6, 2 (2017), 57.
16. Dingqi Yang, Daqing Zhang, Vincent W Zheng, and Zhiyong Yu. 2015. **Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs**. IEEE Transactions on Systems, Man, and Cybernetics: Systems 45, 1 (2015),129–142.
17. Yu Zheng, Yukun Chen, Quannan Li, Xing Xie, and Wei-Ying Ma. 2010. **Understanding transportation modes based on GPS data for web applications**. ACM Transactions on the Web (TWEB)4, 1 (2010), 1–36.
18. Fan Zhou, Qiang Gao, Goce Trajcevski, Kunpeng Zhang, Ting Zhong, and Fengli Zhang. 2018. **Trajectory-user Linking via Variational Autoencoder**. In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18). AAAI Press, 3212–3218.

UNIVERSIDADE FEDERAL
DE SANTA CATARINA