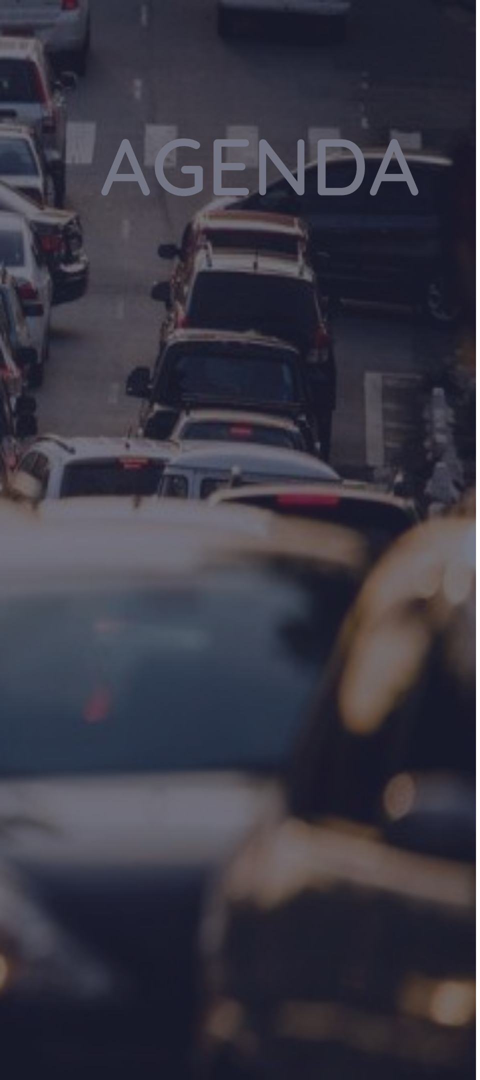
LOCATION PREDICTION ON SYMBOLIC TRAJECTORIES OVER ROAD-NETWORKS

LÍVIA ALMADA CRUZ
SUPERVISOR: Dr. JOSÉ A. F. MACEDO
CO-SUPERVISOR: Dr. KARINE ZEITOUNI



AGENDA

1. INTRODUCTION
2. STATE OF THE ART
3. PROBLEM STATEMENT
4. INITIAL EVALUATION
5. ONGOING WORK
6. CONCLUSION AND NEXT STEPS
7. REFERENCES

c

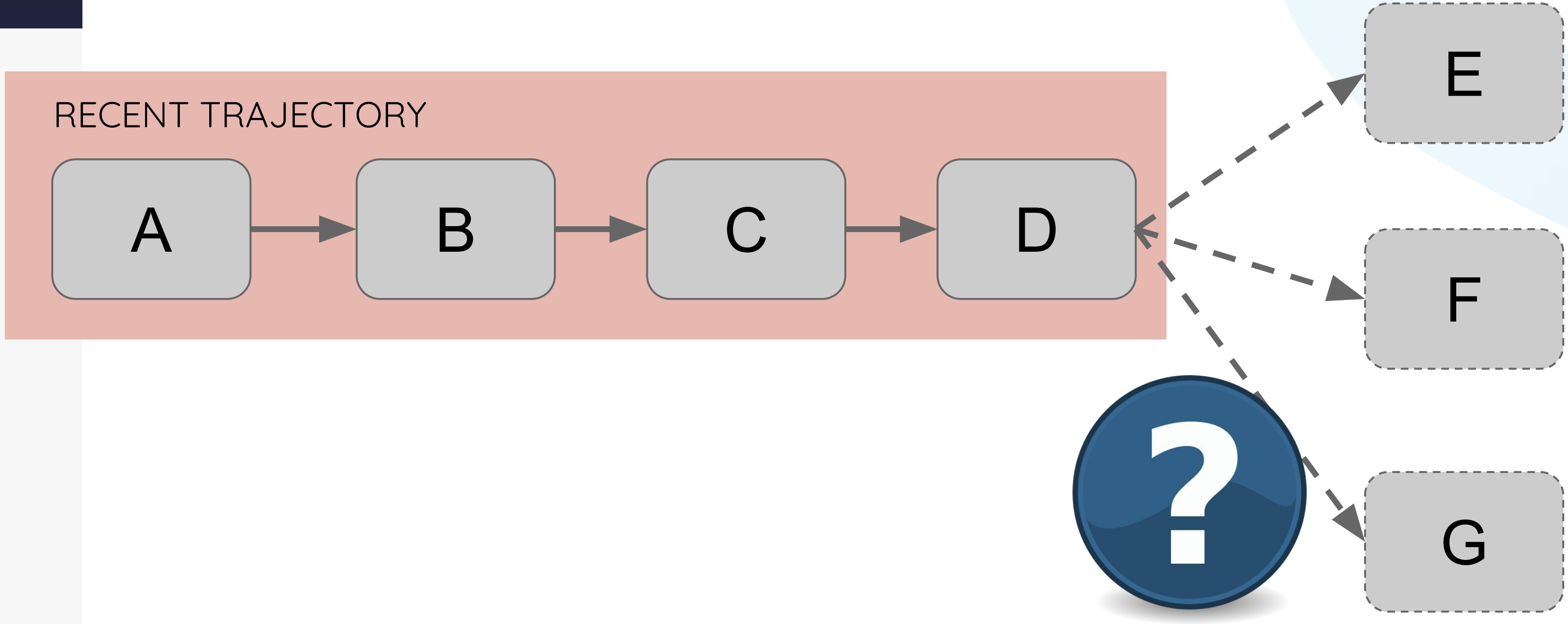
1. INTRODUCTION

What and why location prediction?

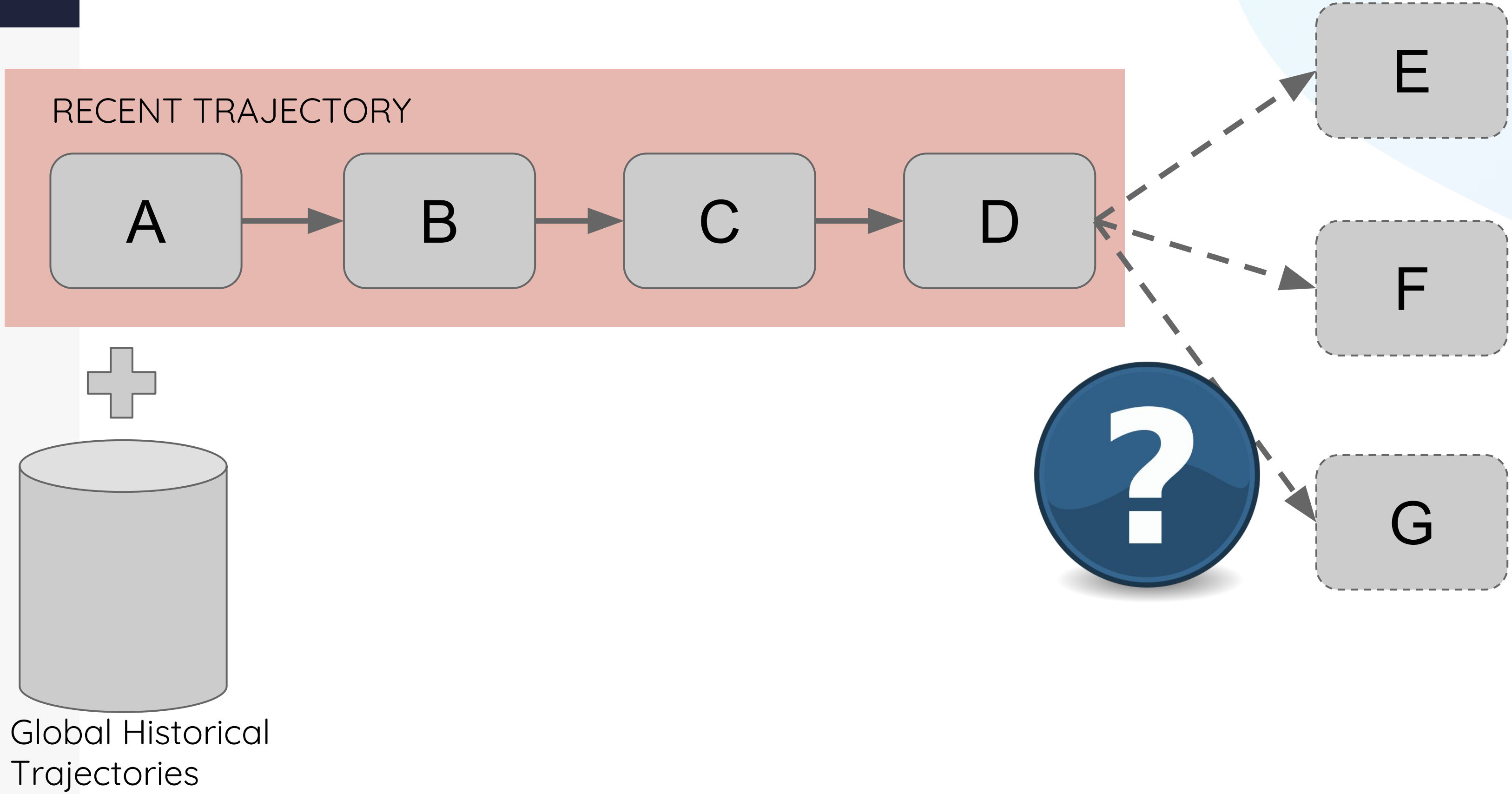
TRAJECTORY DATA

Location-based services, GPS, sensors loops, cameras of traffic surveillance, bluetooth, social media and others

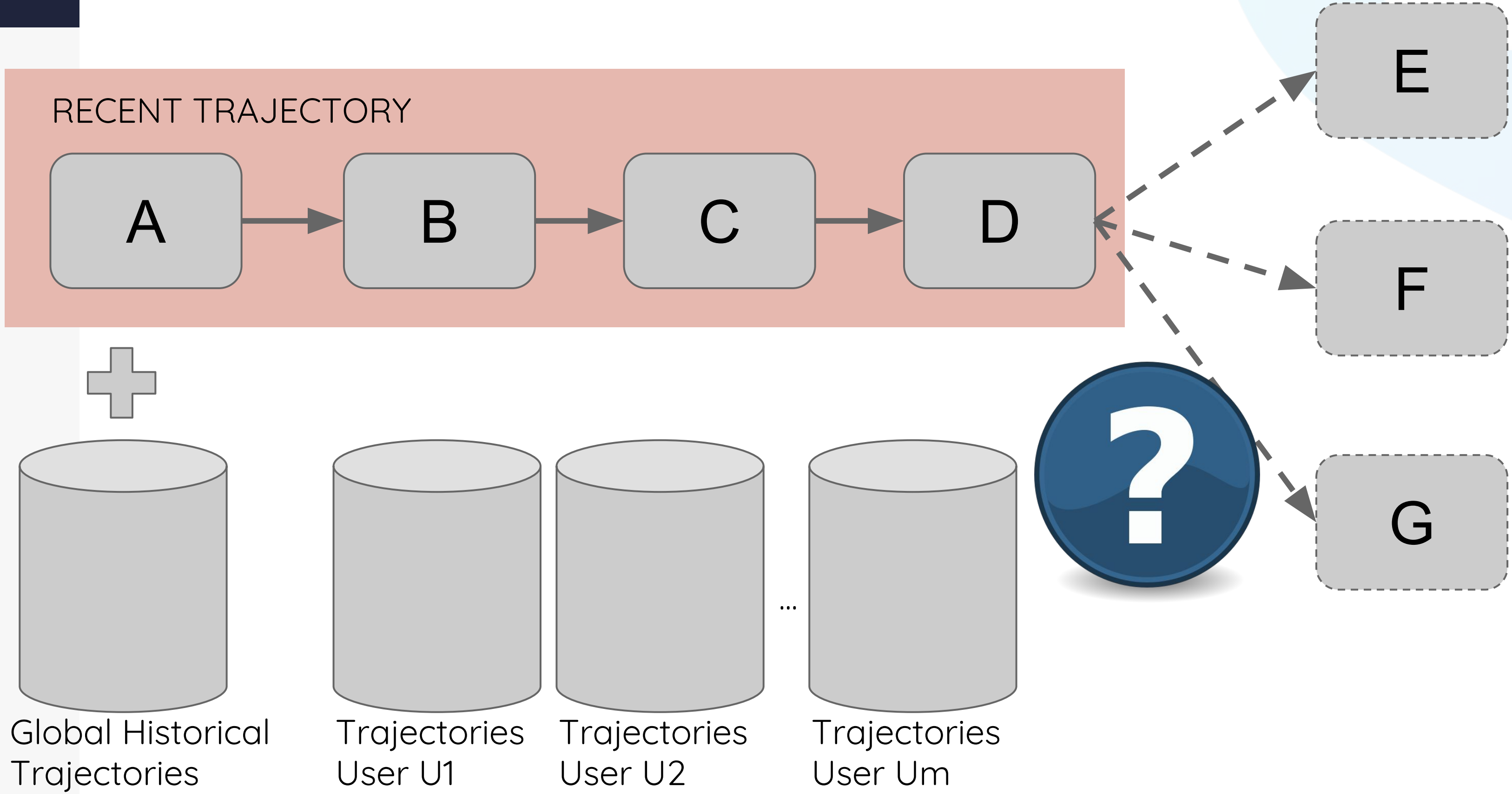
LOCATION PREDICTION



LOCATION PREDICTION



LOCATION PREDICTION



Why location prediction?

Smart Transportation

Urban Planning

(Personalized) Recommender Systems

Resources management

93%

is the potential predictability of human mobility.

Results found by SONG et al. (2010)

CHALLENGES IN LOCATION PREDICTION (FENG et. al., 2018)

Complex sequential transition (time-dependent and high-order)

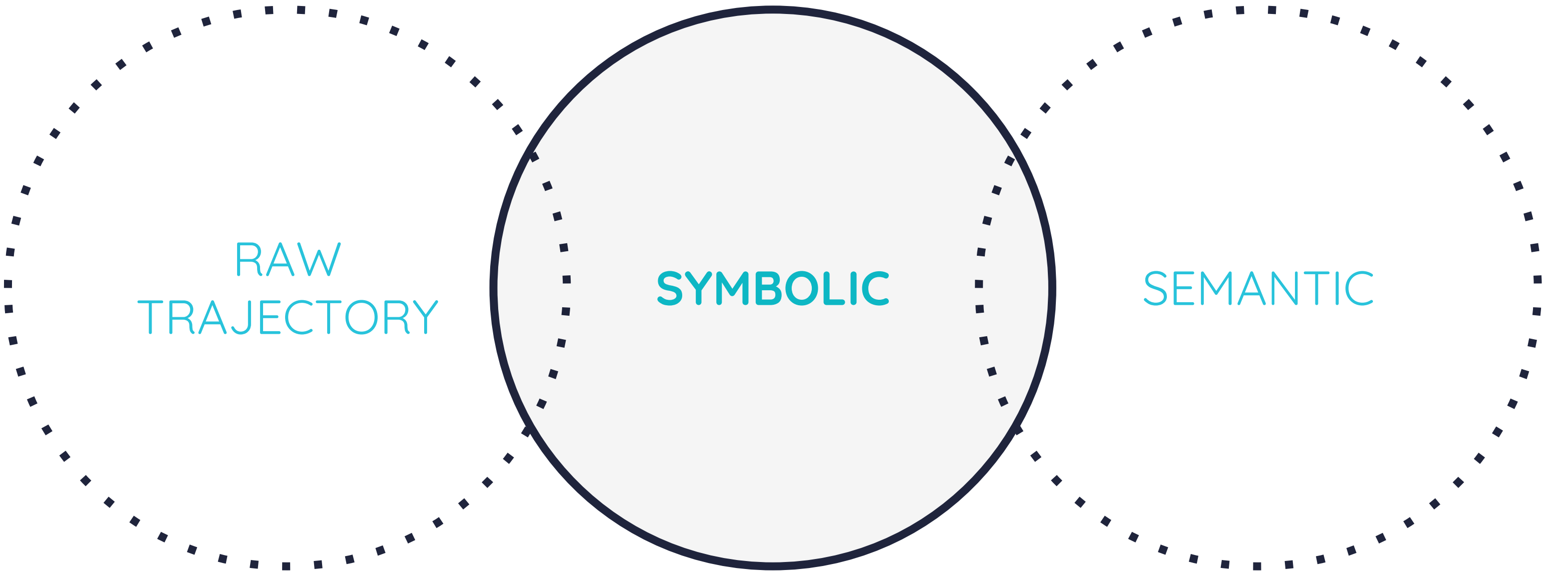
Periodicity in multi-levels

Heterogeneity and sparsity

c

2. STATE OF THE ART

STATE OF THE ART
Trajectory Data



STATE OF THE ART
Raw Trajectory

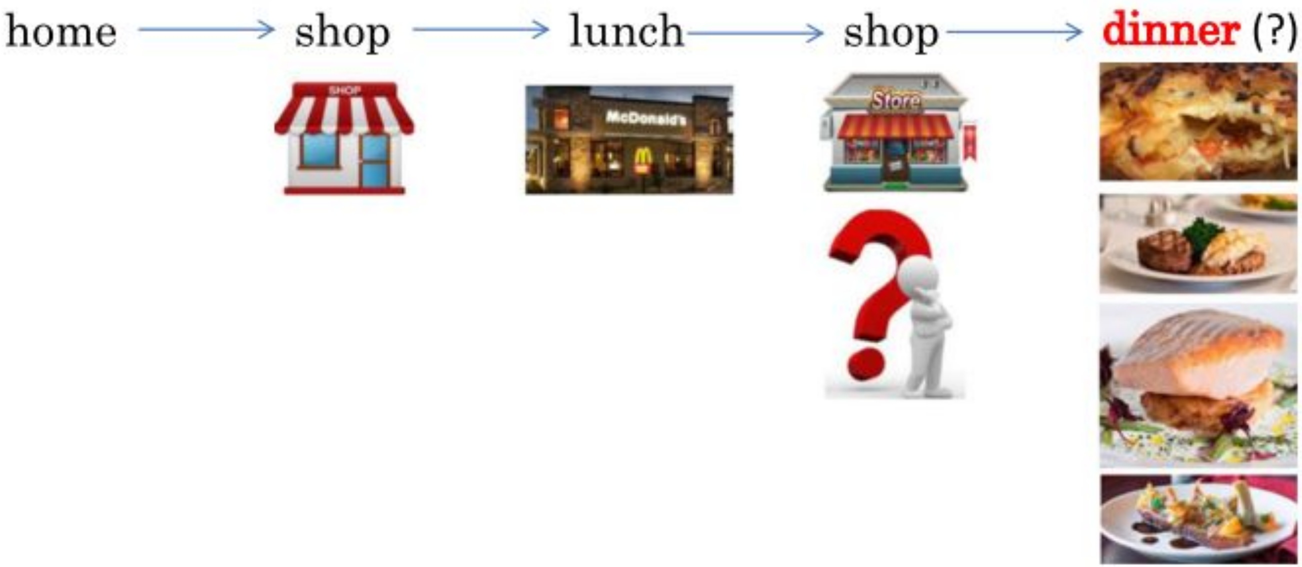
- Sequence of **positions** and **time**
  - {<lat1,lon1,t1>, <lat2,lon2,t2>, <lat3,lon3,t3>, …, <latN,lonN,tN>}
  - Usually restricted to some subset of trajectories
    - e.g. taxi fleet

STATE OF THE ART
Symbolic Trajectory

- Sequence of pairs of locations **labels** (from an alphabet) and **time**
  - {<label1, t1>, <label2, t2>, <label3, time>, …., <labelN, tN>}

STATE OF THE ART
Symbolic Trajectory

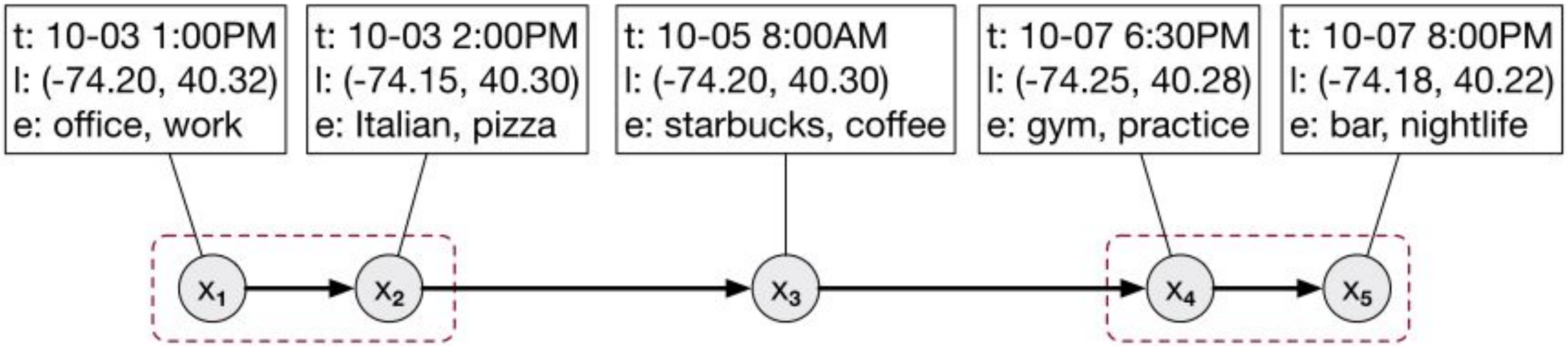
- e.g. Sequence of check-ins from Social Media



Zhao et al. (2017). A time-aware trajectory embedding model for next-location recommendation. Knowledge and Information Systems.

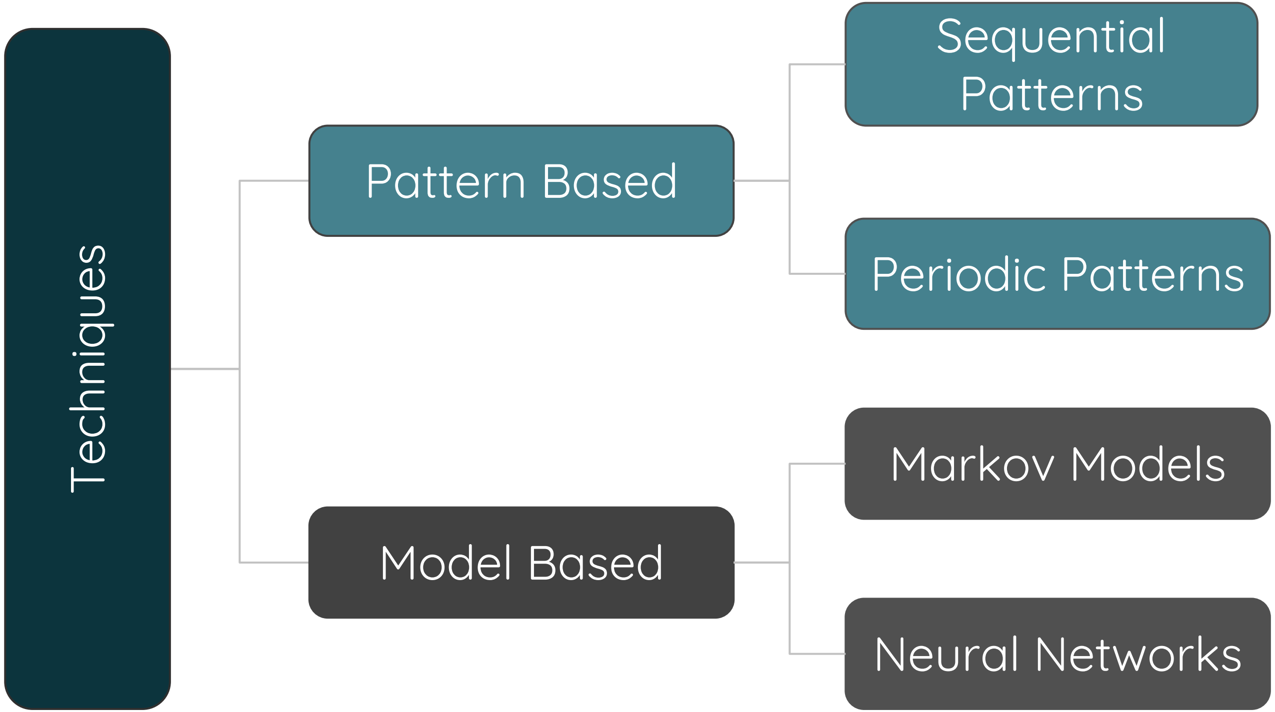
STATE OF THE ART
Semantic Trajectory

- Sequence of pairs of **location** and **time**, plus semantic information
  - e.g. Sequence of check-ins from Social Media plus hashtags



Zhang, C., Zhang, K., Yuan, Q., Zhang, L., Hanratty, T., & Han, J. (2016). GMove : Group-Level Mobility Modeling Using Geo-Tagged Social Media. KDD.

STATE OF THE ART
Techniques



STATE OF THE ART
Pattern-based approaches

- TPRED

  ROCHA et al. (2016). TPRED: a Spatio-Temporal Location Predictor Framework. *IDEAS*.

  - GPS traces
  - Identify stops and builds a suffix tree with transitions between stops
- MyWay

  TRASARTI et al. (2017). MyWay: Location prediction via mobility profiling. Information Systems.

  - GPS traces
  - Clustering to create profiles
  - Matching based on trajectory and profiles geometries

STATE OF THE ART
Pattern-based approaches

- NASARIAN et al. (2018)

  Personalized location prediction for group travellers from spatial-temporal trajectories. Future Generation Computer Systems.

  - Symbolic trajectories + profile information
  - Clustering to identify types of groups using profiles
  - Sequential rules and matching
- CPT+

  Gueniche, T., Fournier-Viger, P., Raman, R., & Tseng, V. S. (2015). CPT+: Decreasing the time/space complexity of the compact prediction tree. Lecture Notes in Computer Science.

  - Prediction tree for sequence prediction
  - Can be applied for symbolic trajectories
  - Does not consider time

STATE OF THE ART
Model-based approaches

Markov based

- GMove
- Zhang et al. (2016). GMove : Group-Level Mobility Modeling Using Geo-Tagged Social Media. KDD.
  - Social media + keywords
  - Embedded HMM
  - Keywords used on group identification

STATE OF THE ART
Model-based approaches

Recurrent Neural Networks

Usually enrich embedded models with information about time, semantic, user preferences

- ST-RNN
  Liu et al. (2016). Predicting the Next Location : A Recurrent Model with Spatial and Temporal Contexts. AAAI..
- SERM
  Yao et al. (2017). SERM: A Recurrent Model for Next Location Prediction in Semantic Trajectories. CIKM.
- TA-TEM
  Zhao et al. (2017). A time-aware trajectory embedding model for next-location recommendation. Knowledge and Information Systems.
- DeepMove
  Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, D. J. (2018). DeepMove : Predicting Human Mobility with Attentional Recurrent Networks. WWW.

STATE OF THE ART
Overview

- More recent approaches focus on models based on Recurrent Neural Networks
- Data sets are restricted to an **specific sample** of trajectories
    - **Raw** trajectories: Particular types of fleet
    - **Symbolic** trajectories: Check-ins/Events
    - **Semantic** trajectories: Check-ins + Textual information
- Do not match the trajectory with the information provided by the road network

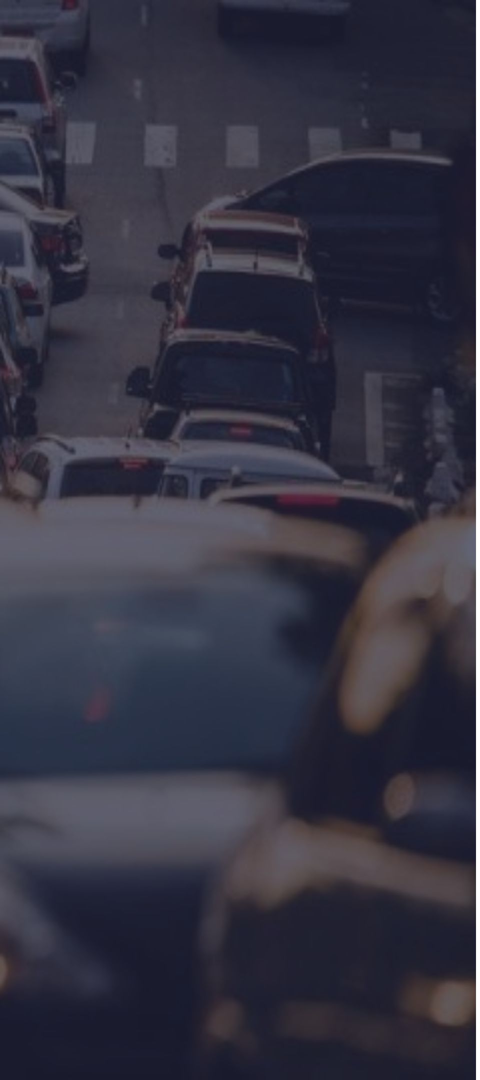
c

3. PROBLEM STATEMENT

Location Prediction for Symbolic Trajectories over the Road-Network

Location Prediction for Symbolic Trajectories over Road-Networks

- **Given:** A set of symbolic trajectories over a road-network
  - Trajectories are a sequence of **landmarks** over the road-network
- **Goal**: Predict the next location given the recent trajectory
- More **complete** dataset of symbolic trajectories
  - More than one million of vehicles
  - **Exhaustive** types of trajectories
- Consider **network** characteristics + **Time** + **User** preferences



General Goals

Predict next location on symbolic trajectories considering

- Road-network
  - Enrich the model with network information
- Trajectory sparsity
- Temporal context
- Big data volume

c

4. INITIAL EVALUATION

INITIAL EVALUATION
Goals

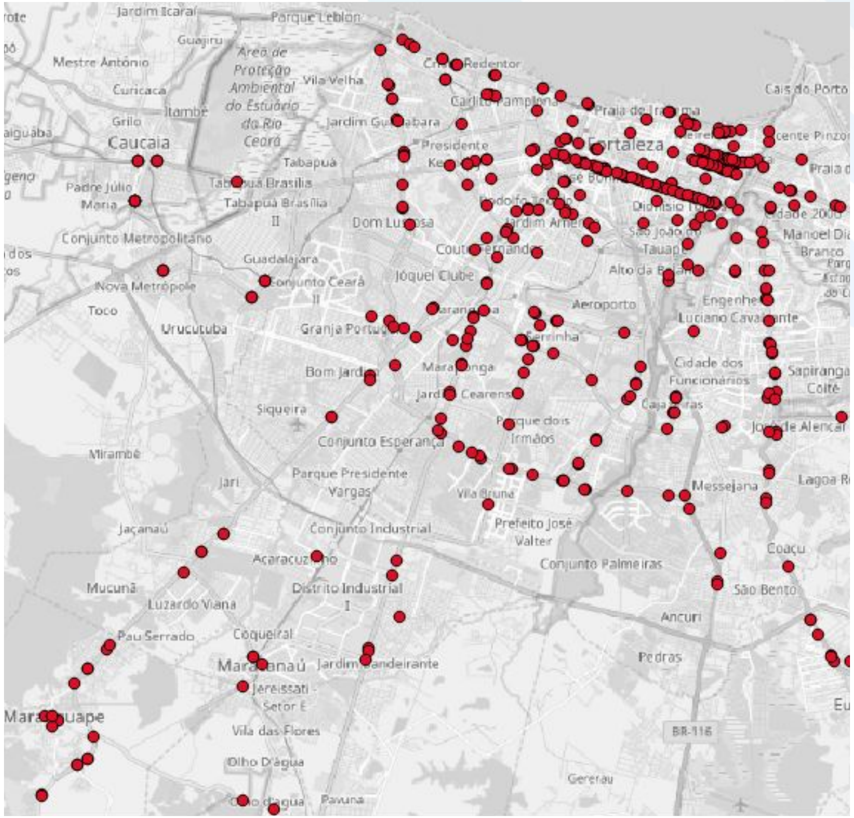
- Understand the dataset
- Apply more basic existing approaches for sequential prediction to see how it works
- Question: Is the sequence prediction enough to predict the trajectories without time?

INITIAL EVALUATION
Dataset

- Landmarks:

  746 cameras and sensors of traffic surveillance

- Main collected features:
  - timestamp
  - position
  - plate (= vehicle id)
  - speed
  - type of road
  - and others



INITIAL EVALUATION
Dataset

Whole dataset:

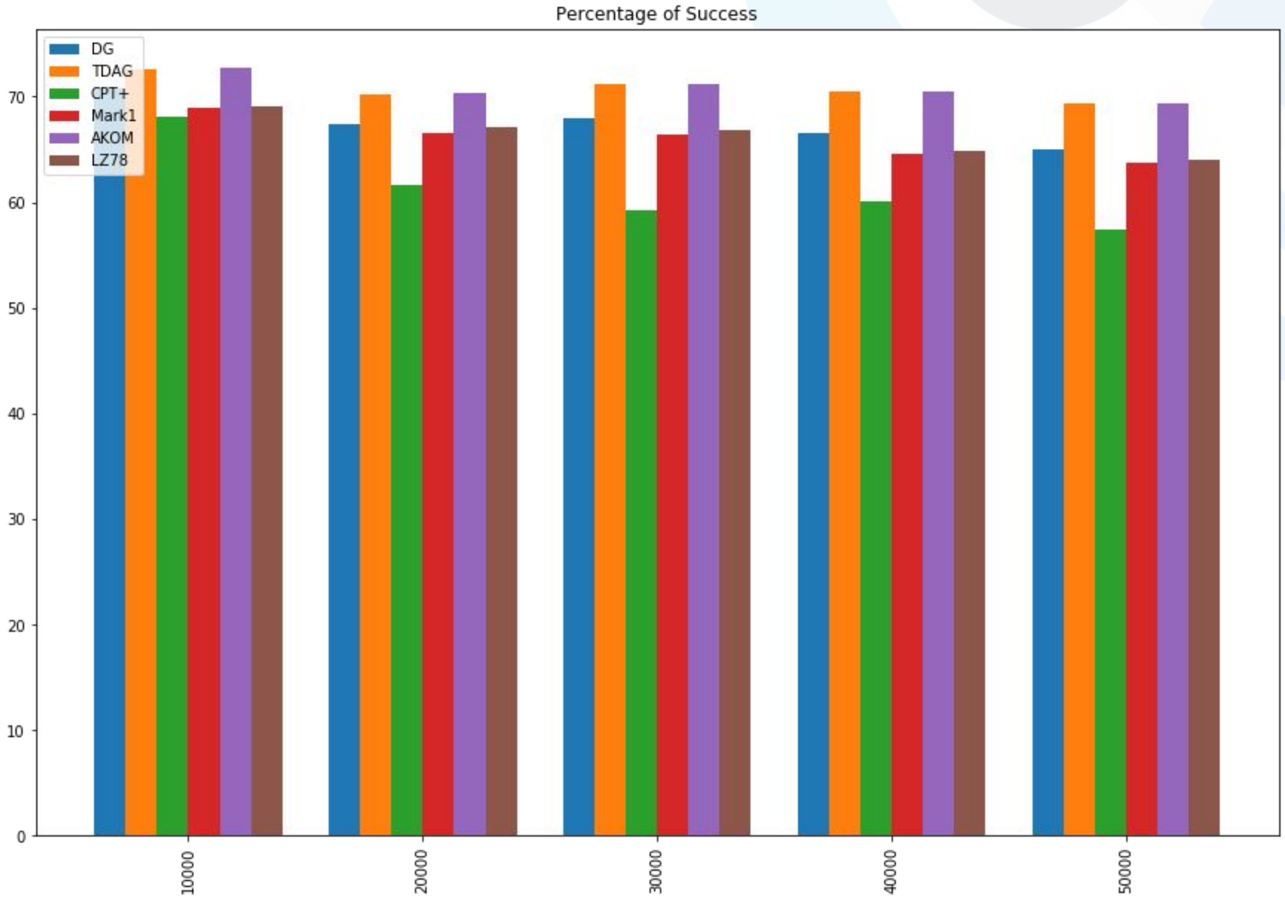
- More than one million of distinct vehicles
- Total number of points: 367,857,722
- From 2017-04-04 to 2018-05-01

INITIAL EVALUATION
Analysis of Predictability with Sequence Predictors

- **Algorithms for Sequence Prediction**
  - DG , TDAG, CPT+, Markov Model First Order, Markov Model Kth Order (AKOM), LZ78
- **All algorithms available on the SPMF Library**
  - Size of window = 3
  - Consequent size = 2 (To next locations are the ground truth)
- **From 10,000 to 50,000 trajectories with size ≥ 10**

INITIAL EVALUATION
Success

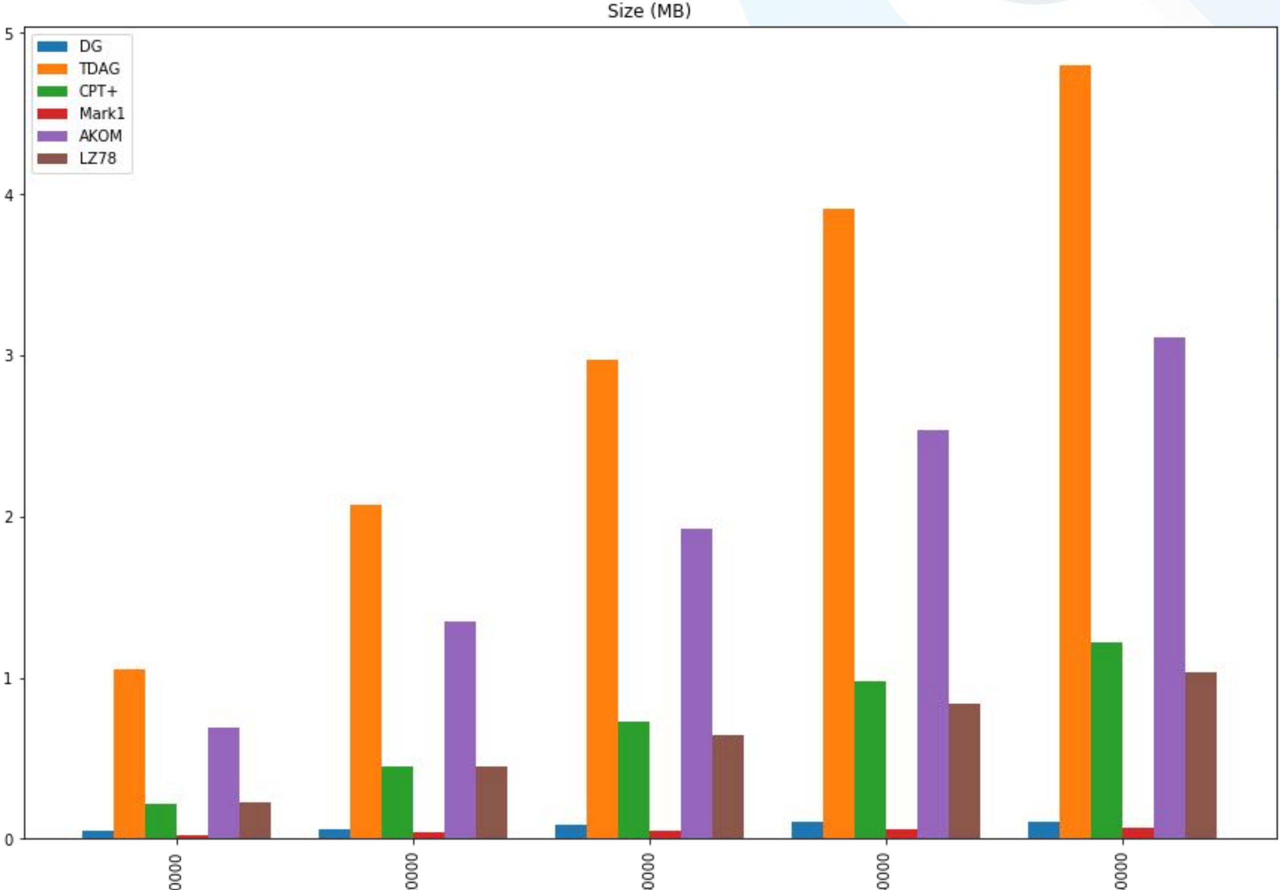
	DG	TDAG	CPT+	Mark1	AKOM	LZ78
10000	70.710	72.542	68.077	68.944	72.776	69.044
20000	67.355	70.222	61.656	66.589	70.405	67.055
30000	67.948	71.137	59.171	66.459	71.237	66.848
40000	66.586	70.436	60.128	64.645	70.527	64.895
50000	65.049	69.295	57.483	63.736	69.369	64.089



- From 60% to 70% of accuracy
- **not temporal**
- **only global**

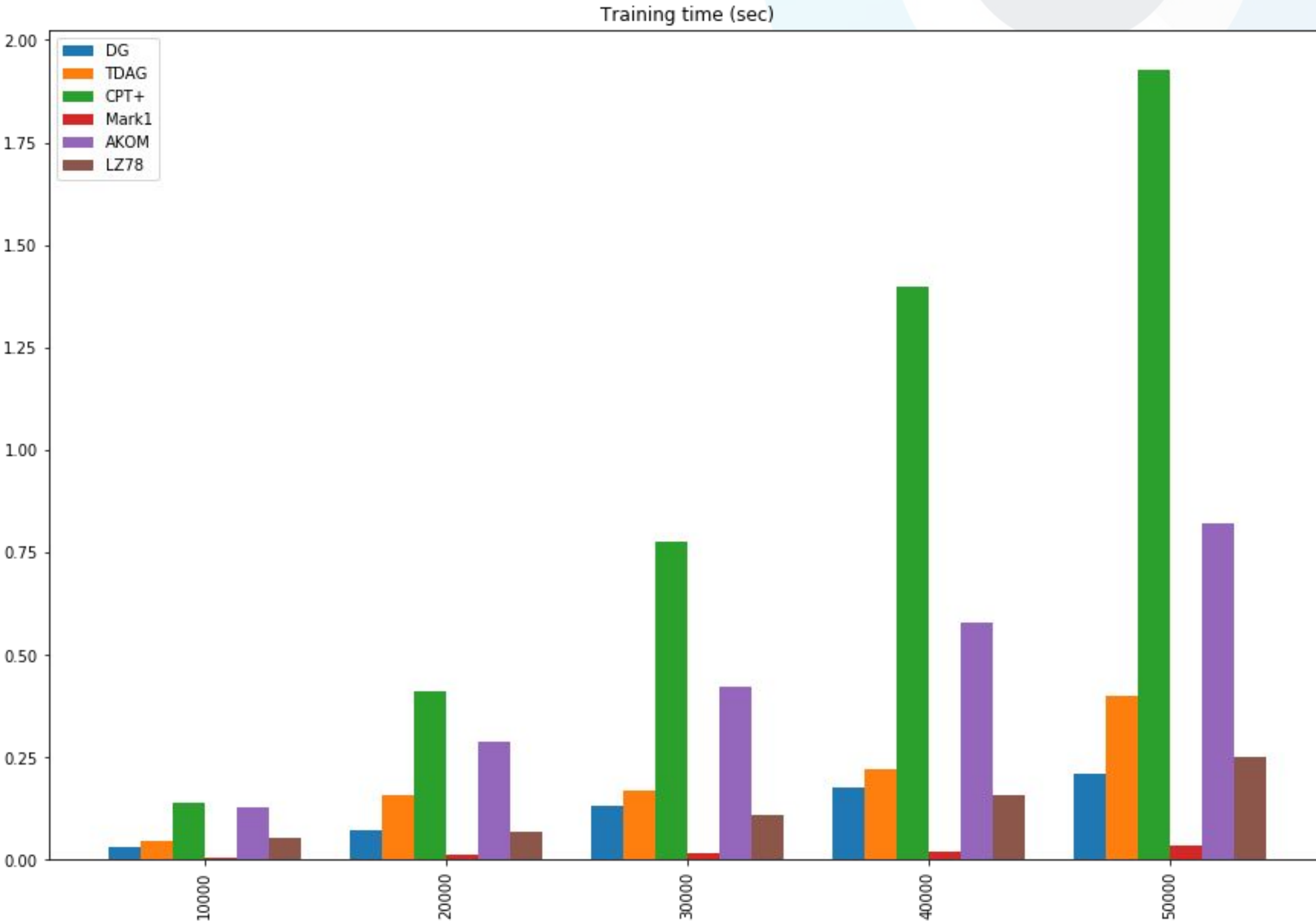
INITIAL EVALUATION
Memory Consumption

	DG	TDAG	CPT+	Mark1	AKOM	LZ78
10000	0.051	1.055	0.216	0.027	0.688	0.222
20000	0.060	2.076	0.445	0.037	1.345	0.446
30000	0.090	2.971	0.728	0.051	1.925	0.644
40000	0.106	3.908	0.976	0.064	2.533	0.838
50000	0.110	4.803	1.216	0.072	3.111	1.032



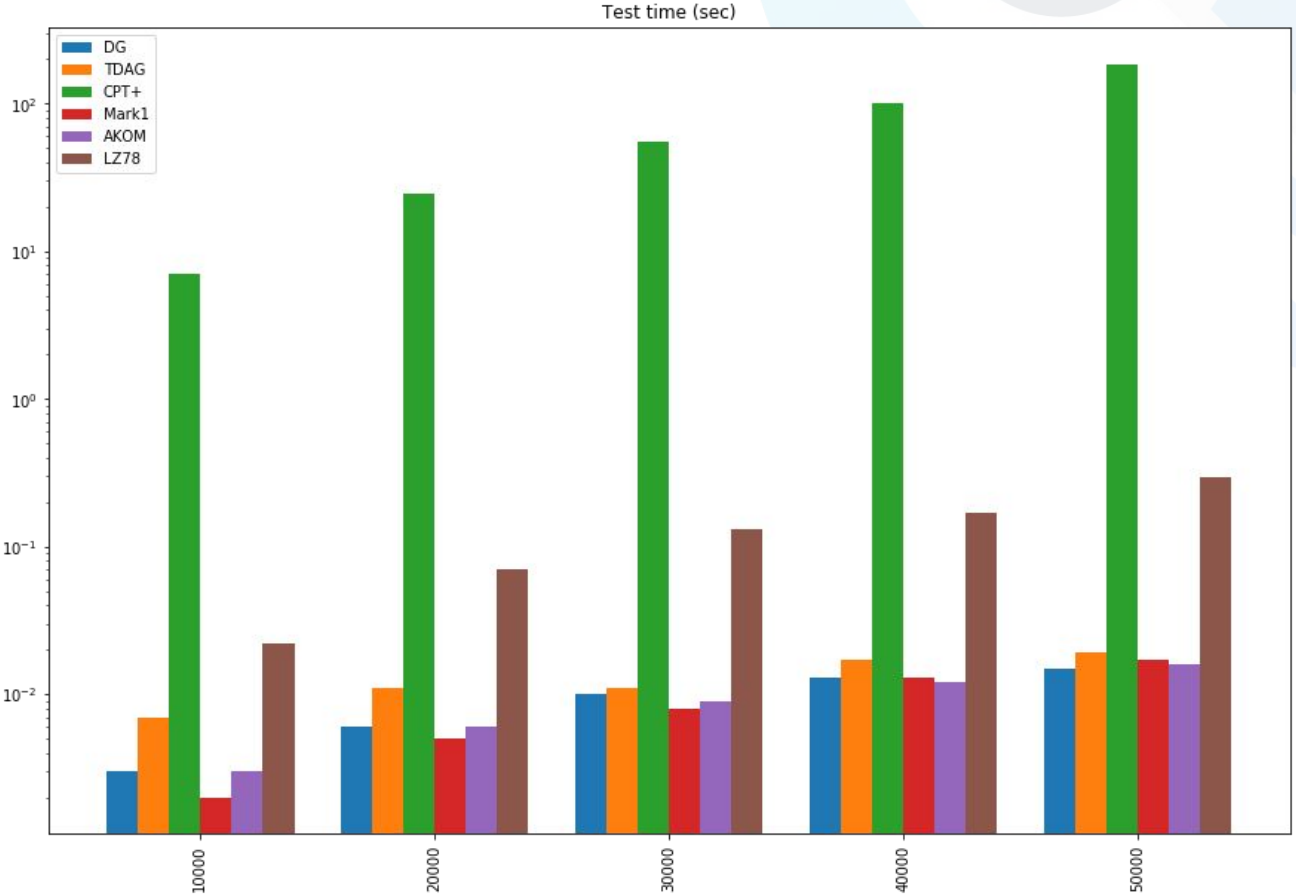
INITIAL EVALUATION
Training Time

	DG	TDAG	CPT+	Mark1	AKOM	LZ78
10000	0.030	0.047	0.141	0.005	0.130	0.054
20000	0.072	0.159	0.410	0.011	0.290	0.068
30000	0.133	0.170	0.775	0.018	0.422	0.109
40000	0.175	0.221	1.398	0.022	0.578	0.157
50000	0.209	0.400	1.928	0.035	0.821	0.251



INITIAL EVALUATION
Test Time

	DG	TDAG	CPT+	Mark1	AKOM	LZ78
10000	0.003	0.007	7.025	0.002	0.003	0.022
20000	0.006	0.011	24.650	0.005	0.006	0.070
30000	0.010	0.011	55.028	0.008	0.009	0.131
40000	0.013	0.017	101.834	0.013	0.012	0.171
50000	0.015	0.019	185.749	0.017	0.016	0.294



INITIAL EVALUATION
Results

- Models for sequence prediction able to predict with accuracy from 60% to 70%
- Does not take into account:
  - Temporal context of trajectories
  - Road Network
  - User preferences

EXPLORING WORD2VEC MODEL

- **Word Embedding**
  - Represent an item (word) as a vector of numbers
  - Between two words depends on the vector representation
- **Ex. One hot representation**
  - High dimensional

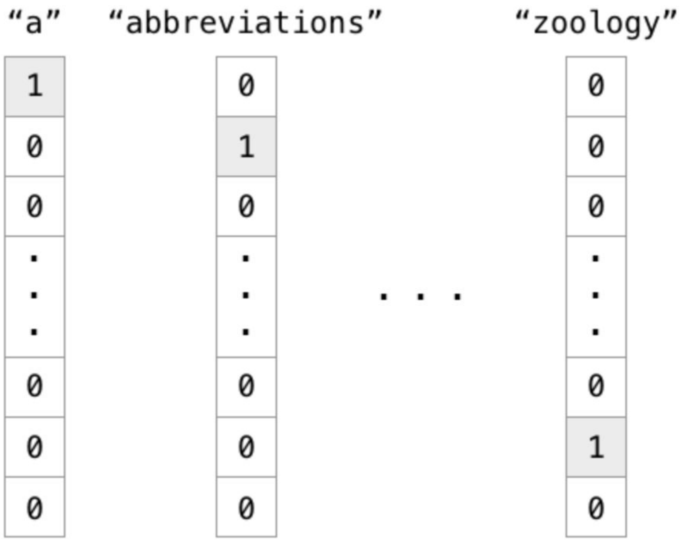


Figure. One hot representation.

EXPLORING WORD2VEC MODEL

- **Word2Vec**
  Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. NIPS.
  - embedded model
  - predictor based on neural networks
  - lower dimensional vector
  - context of words

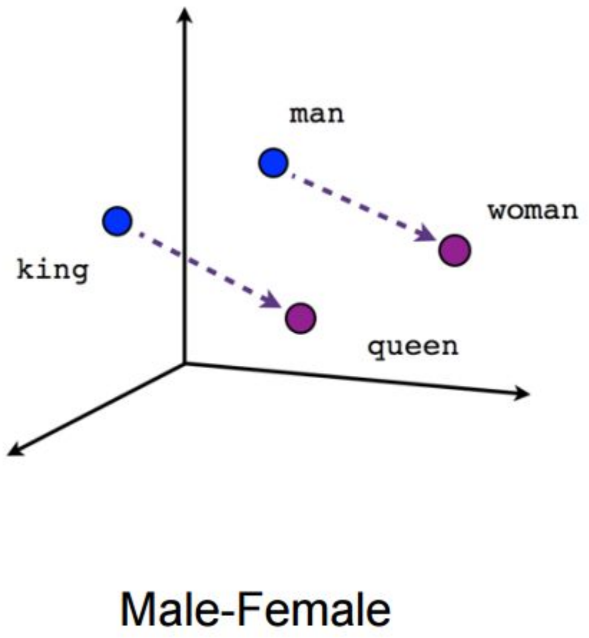


Figure. Vizualize Word2Vec.
https://www.tensorflow.org/tutorials/word2vec

EXPLORING WORD2VEC MODEL

Capture links between sensors sharing trajectories

- **DATA SET: 400 SENSORS**
- **ONE WEEK OF OBSERVATIONS**

Figure. Links between KNN according to Euclidean Distance (k=4).

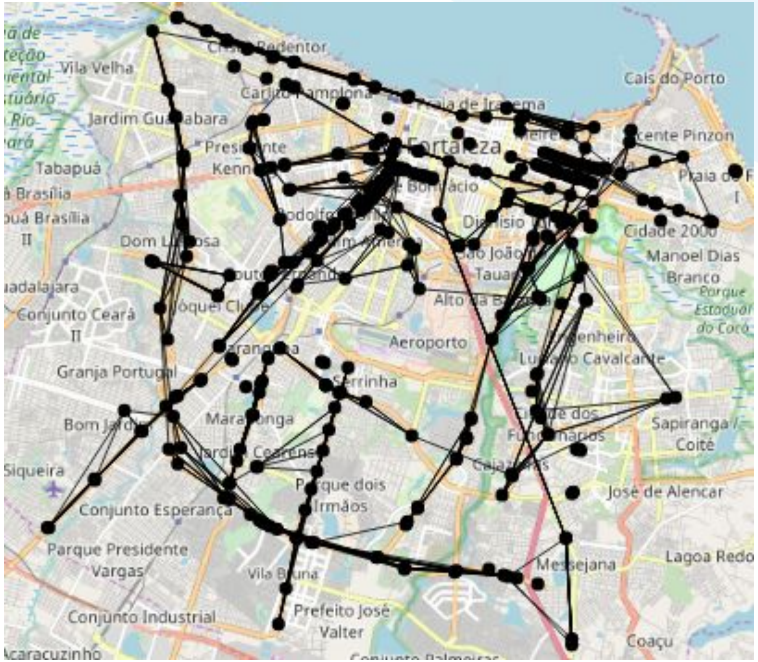
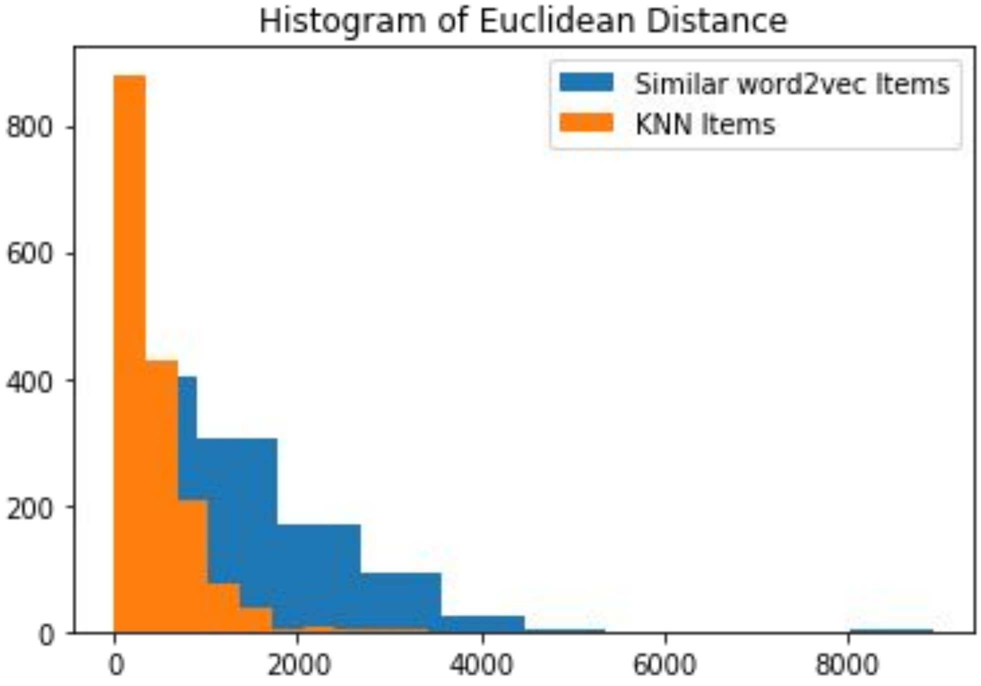


Figure. Links between KNN according to word2vec similarity. (k=4)

EXPLORING WORD2VEC MODEL



c

5. ONGOING WORK

RECURRENT NEURAL NETWORKS

- RNN are the state-of-art for location prediction
- "Memory" which captures previous information
- LSTM (Long Short Term Memory) for long sequences

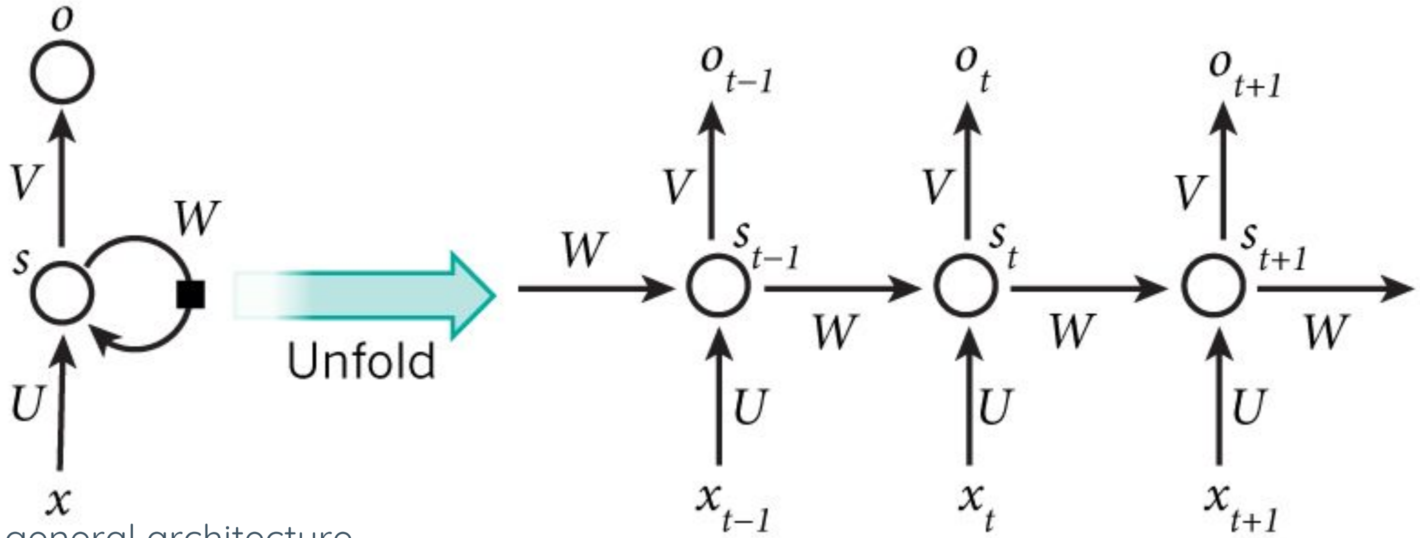


Figure. RNN general architecture.
http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

RECURRENT NEURAL NETWORKS
State of art for location prediction

- ST-RNN (LIU et al. 2016)
  - Check-ins/Events
  - Continuous values in spatial and temporal contexts
- SERM (YAO et al. 2017)
  - Check-ins + time + text messages
- TA-TEM (ZHAO et al. 2017)
  - Check-ins + dynamic and static user preferences + weekly and daily patterns
- DeepMove (FENG et al. 2018)
  - Symbolic trajectories

c

5. CONCLUSION AND NEXT STEPS

CONCLUSION AND NEXT STEPS

State of the art

- RNN are the state-of-art for location prediction
- Restricted dataset vs More completed (Exhaustive types of trajectories)
- Sequential predictors can predict symbolic trajectory, but do not consider time

CONCLUSION AND NEXT STEPS

Next steps

- How to consider the road network to enrich the models and reach better accuracy?
- Explore RNN + Network information

REFERENCES

- Rocha, C. L., Brilhante, I. R., Lettich, F., Andrade, R., & Orlando, S. (2016). TPRED: a Spatio-Temporal Location Predictor Framework. IDEAS.
- Trasarti, R., Guidotti, R., Monreale, A., & Giannotti, F. (2017). MyWay: Location prediction via mobility profiling. Information Systems, 64, 350–367. https://doi.org/10.1016/j.is.2015.11.002
- Song, C., Qu, Z., Blumm, N., & Barabási, A. L. (2010). Limits of predictability in human mobility. Science, 327(5968), 1018–1021. https://doi.org/10.1126/science.1177170
- Zhao, W. X., Zhou, N., Sun, A., Wen, J. R., Han, J., & Chang, E. Y. (2017). A time-aware trajectory embedding model for next-location recommendation. Knowledge and Information Systems, 1–21. https://doi.org/10.1007/s10115-017-1107-4
- Zhang, C., Zhang, K., Yuan, Q., Zhang, L., Hanratty, T., & Han, J. (2016). GMove : Group-Level Mobility Modeling Using Geo-Tagged Social Media. KDD., 1305–1314. https://doi.org/10.1145/2939672.2939793
- Naserian, E., Wang, X., Dahal, K., Wang, Z., & Wang, Z. (2018). Personalized location prediction for group travellers from spatial-temporal trajectories. Future Generation Computer Systems, 83, 278–292. https://doi.org/10.1016/j.future.2018.01.024
- Gueniche, T., Fournier-Viger, P., Raman, R., & Tseng, V. S. (2015). CPT+: Decreasing the time/space complexity of the compact prediction tree. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9078, 625–636. https://doi.org/10.1007/978-3-319-18032-8_49
- Liu, Q., Wu, S., Wang, L., & Tan, T. (2016). Predicting the Next Location : A Recurrent Model with Spatial and Temporal Contexts. Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016), 194–200. https://doi.org/10.1016/j.egypro.2016.11.209
- Zhao, W. X., Zhou, N., Sun, A., Wen, J. R., Han, J., & Chang, E. Y. (2017). A time-aware trajectory embedding model for next-location recommendation. Knowledge and Information Systems, 1–21. https://doi.org/10.1007/s10115-017-1107-4
- Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, D. J. (2018). DeepMove : Predicting Human Mobility with Attentional Recurrent Networks. Www, 2. https://doi.org/10.1145/3178876.3186058
- SPMF: An Open-Source Data Mining Library, http://www.philippe-fournier-viger.com/spmf/

Merci!
Thank you!
Obrigada!

livia.almada@ufc.br