# Representation Learning and Forecasting for Inter-related Time Series


Presented by Jingwei ZUO

DAVID Lab, UVSQ, Université Paris-Saclay

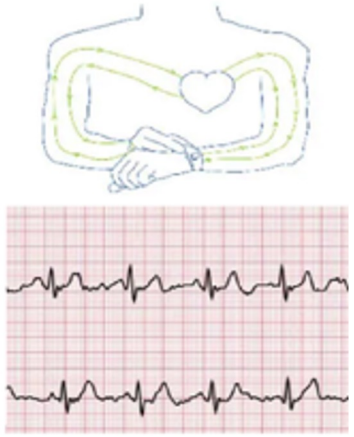

9th may, 2022

**Jury members**

| | | |
|---|---|---|
| Angela BONIFATI | Professeure, Université Claude Bernard Lyon 1 | Rapportrice |
| Engelbert MEPHU NGUIFO | Professeur, Université Clermont Auvergne | Rapporteur |
| Antoine CORNUÉJOLS | Professeur, AgroParisTech | Examinateur |
| Romain TAVENARD | Maitre de conférence (HDR), Université de Rennes 2 | Examinateur |
| | | |
| Karine ZEITOUNI | Professor, UVSQ, Université Paris-Saclay | Directrice de thèse |
| Yehia TAHER | Maitre de conférence, UVSQ, Université Paris-Saclay | Co-encadrant de thèse |

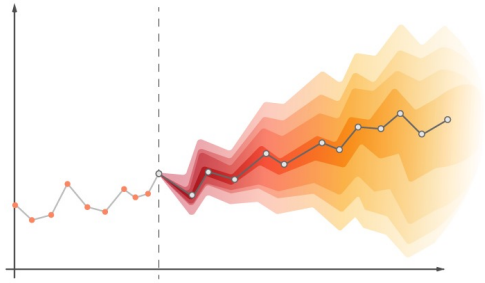
# Time series applications

Electrocardiograph (ECG)
- Arrhythmia?



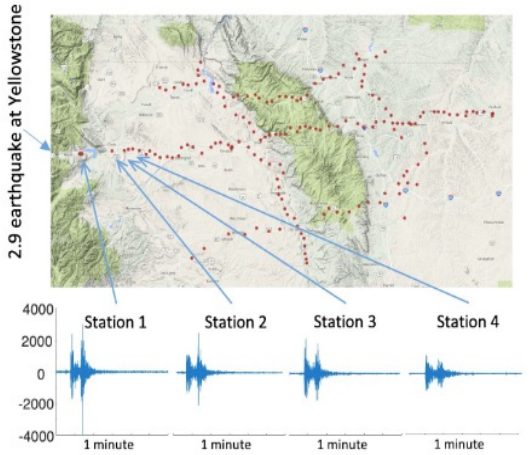
Time series classification

Traffic forecasting
- Traffic flows in the future



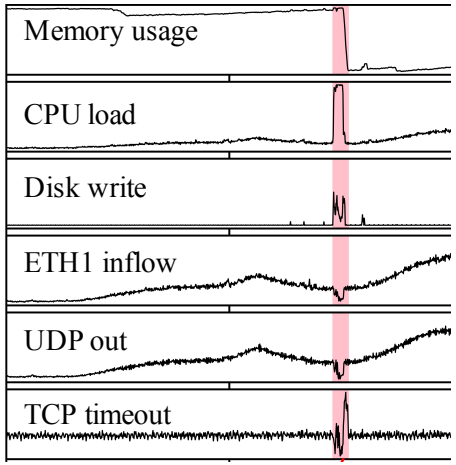
Time series forecasting

Seism
- Earthquake signals



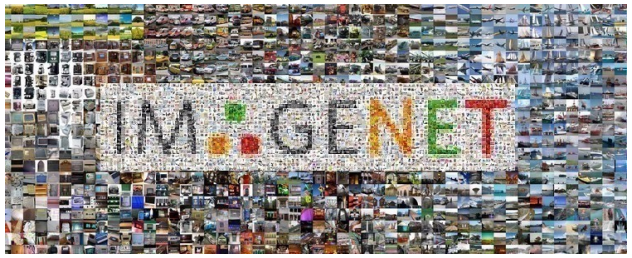

Motif discovery

Server machine
- Abnormal activities

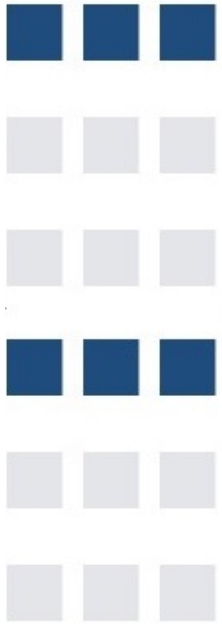



Anomaly detection

# Representation learning




High-dimensional data



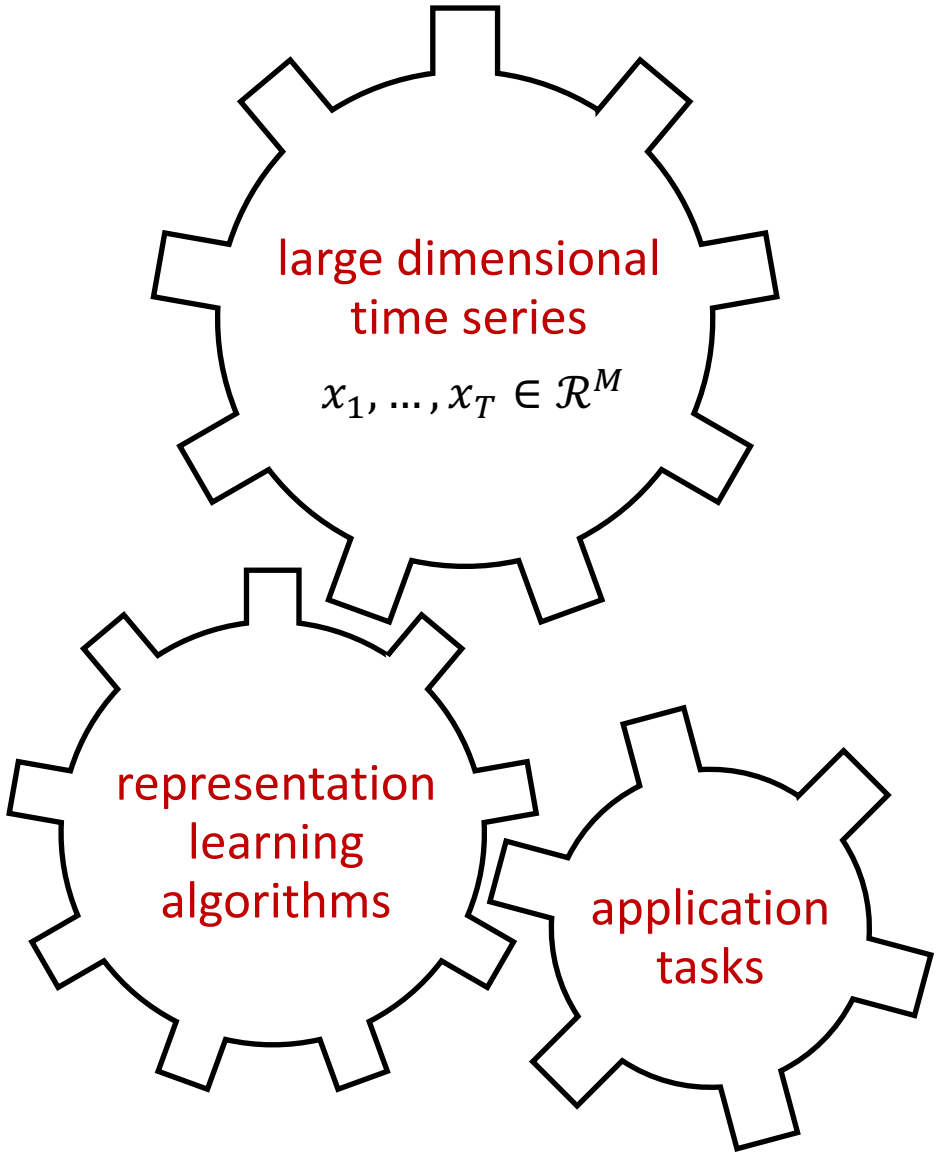
Low-dimensional features

Application tasks

Input data

Intermediate representation

# Why time series representation learning?




- **High** dimensionality
- **Complex** data with inter-relationships
  - temporal relationship
  - inter-variable relationship
- **Various** learning tasks
  - Classification
  - Forecasting
  - Anomaly detection
  - etc.

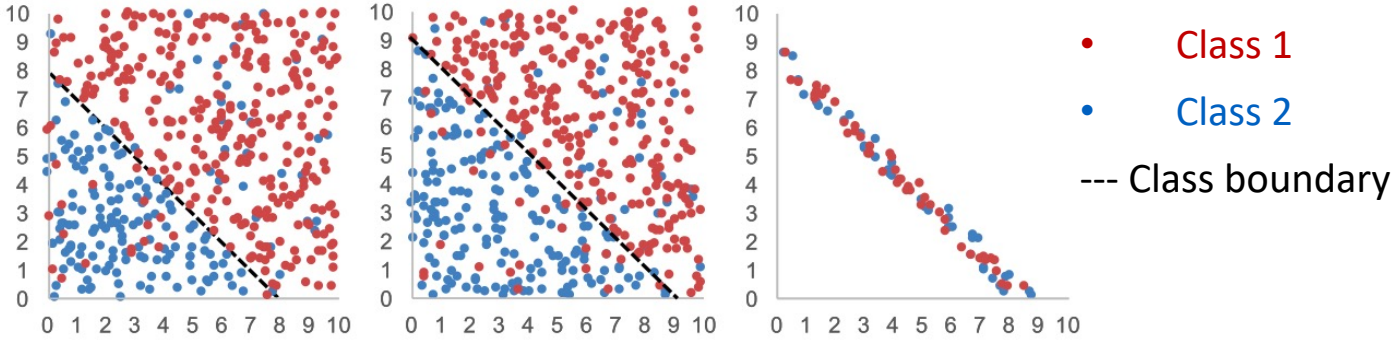
# Challenges in time series representation learning

- **Complex** data with inter-relationships
  - temporal relationships
  - inter-variable relationships
- **Complex** application contexts
  - Streaming context
  - Single source & multiple sources
  - Label shortage
  - Missing values
  - Interpretability & Explainability
  - etc.

➢ No standard representation which fits all the contexts

# Context-related challenges

1. Streaming context

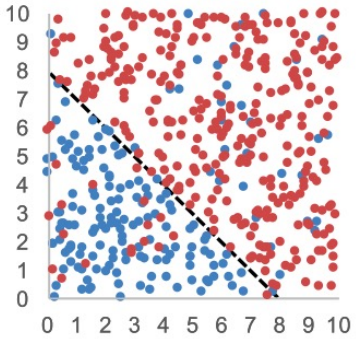



2. Labeling constraint

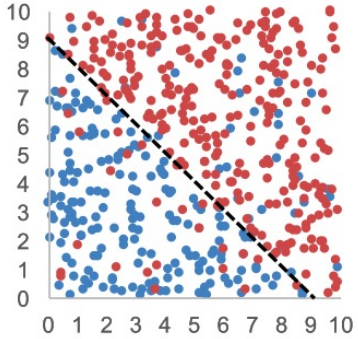

3. Data quality issues, e.g., missing values in Smart City sensor data

# Data complexity challenges

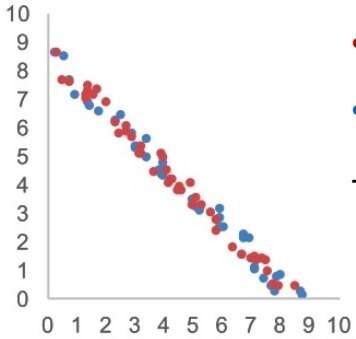
1. Streaming context
   - ➢ Temporal relationships



Data at time t



Data at time t+1

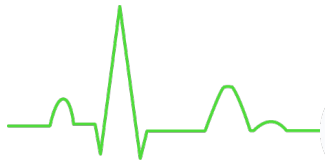


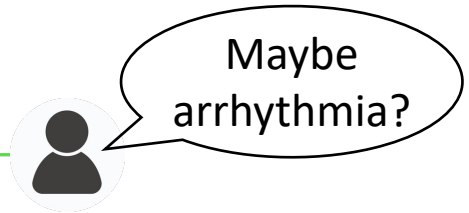
Data in drift region

- Class 1
- Class 2

--- Class boundary

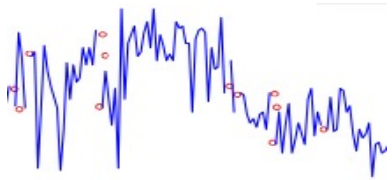
2. Labeling constraint
   - ➢ Temporal & Inter-variable relationships

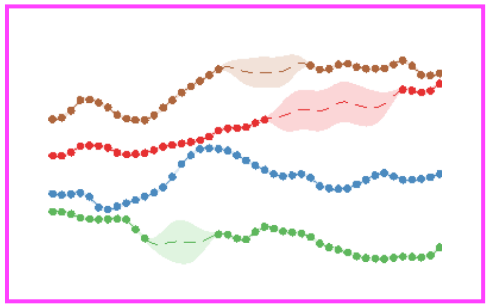







3. Data quality issues, e.g., missing values in Smart City sensor data
   - ➢ Temporal & Inter-variable relationships





# Our contributions

- Streaming context
- Temporal relationships

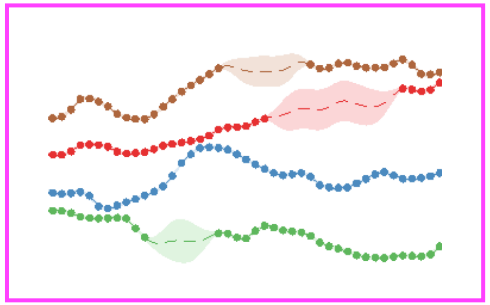
C1: Dynamic feature learning from time series stream

- Labeling constraint
- Temporal & Inter-variable relationships

C2: Semi-supervised representation learning from multivariate time series


- Data quality issues, e.g., missing values in Smart City sensor data
- Temporal & Inter-variable relationships

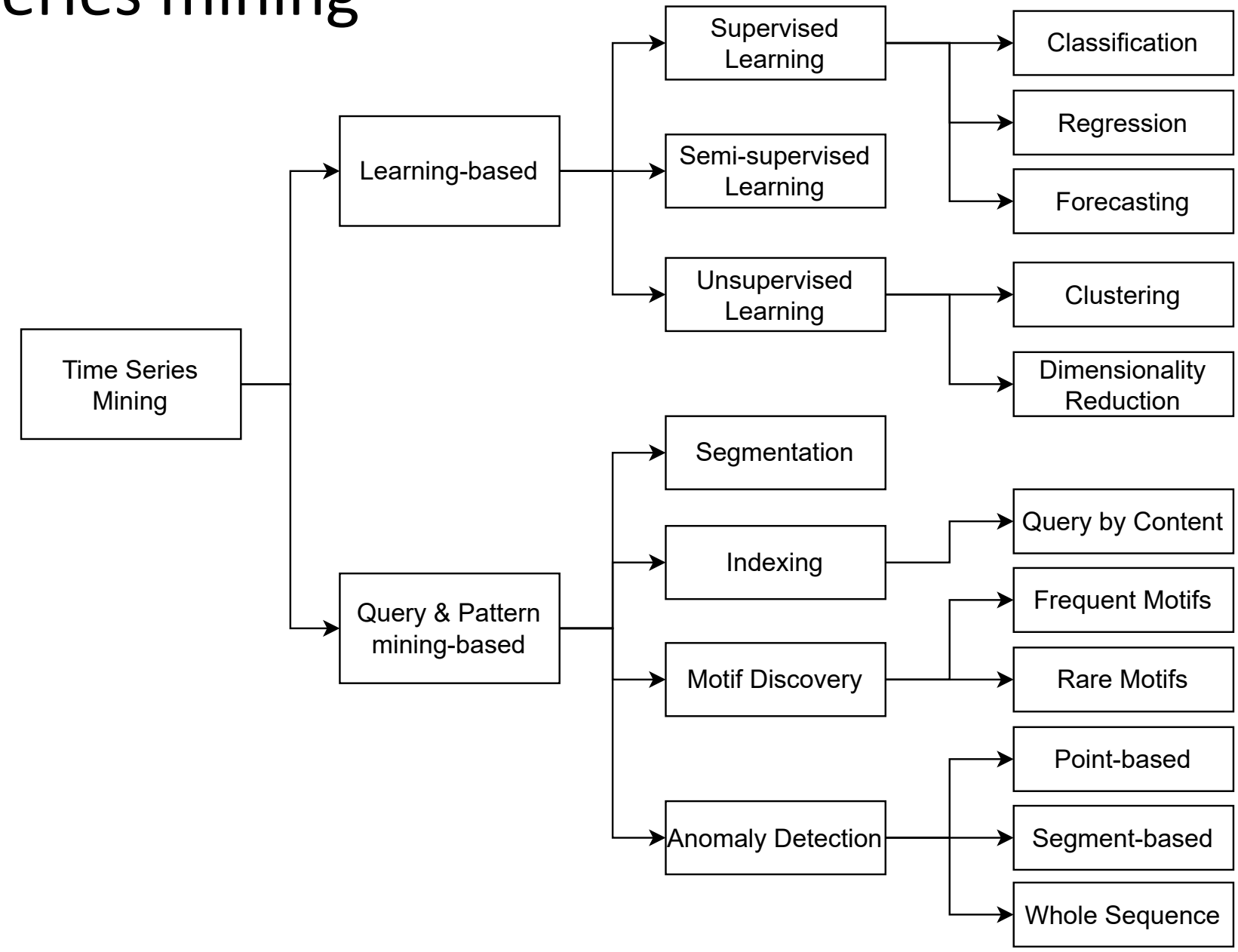
C3: Geo-located time series forecasting with missing values



# Outline

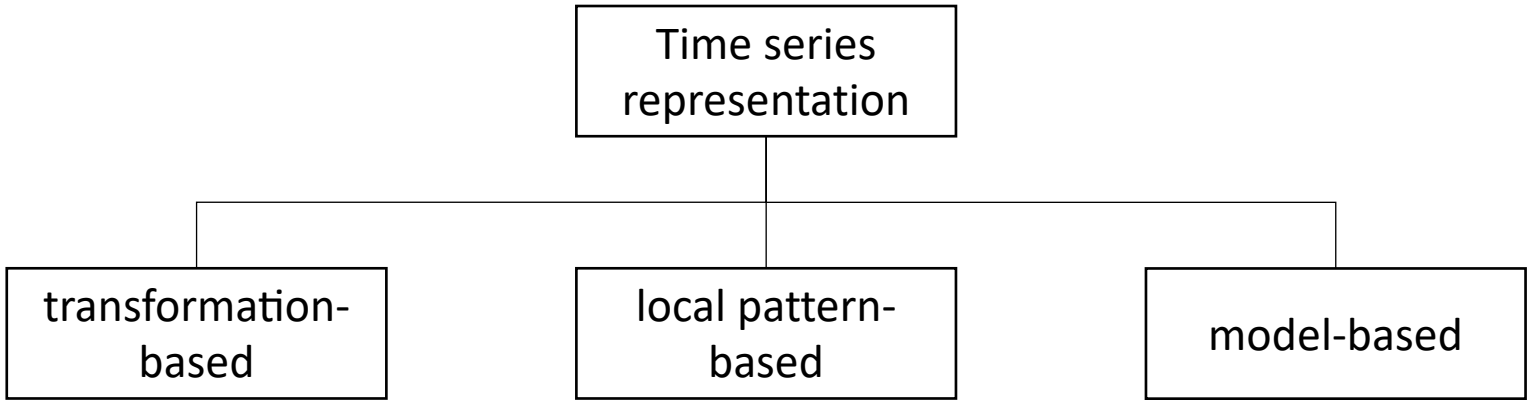
- Introduction
- Background
  - Time series mining
  - Time series representation
- ISMAP: Dynamic Feature Learning on Time Series Stream
- SMATE: Semi-supervised Learning on Multivariate Time Series
- GCN-M: Geo-located Time Series Forecasting with Missing Values
- Conclusion and perspectives

# Time series mining

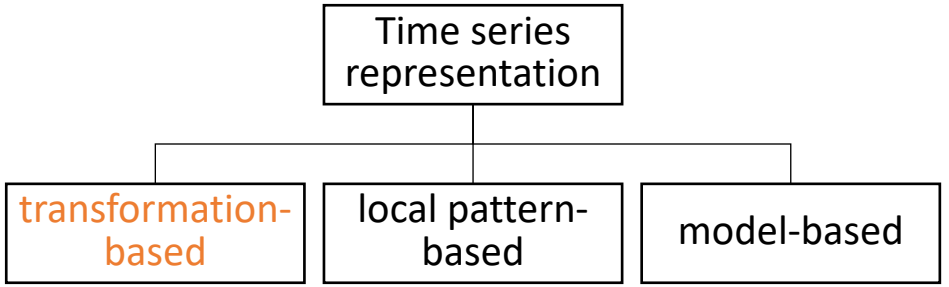


# Time series representation

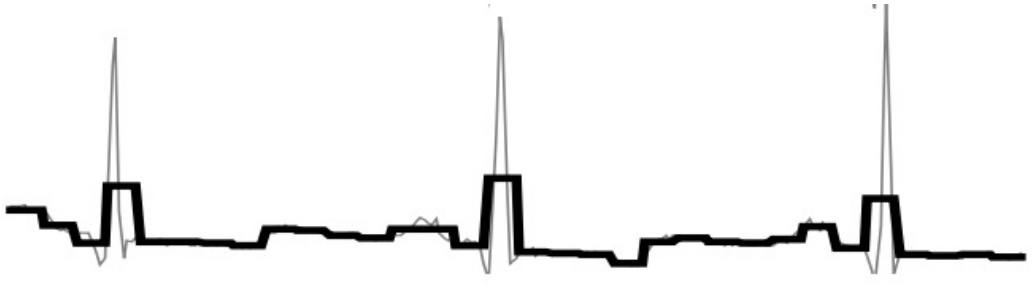
- Time series $\mathrm{x} \in \mathcal{R}^{T \times M}$
- Time series representation $\mathrm{r} \in \mathcal{R}^{T' \times M'}$ : a summarized feature set which accurately describes $\mathrm{x}$
  - $T' \times M' < T \times M$
  - Minimize $\mathrm{Loss}(\mathrm{x}, \mathrm{r})$
- Different types of representations



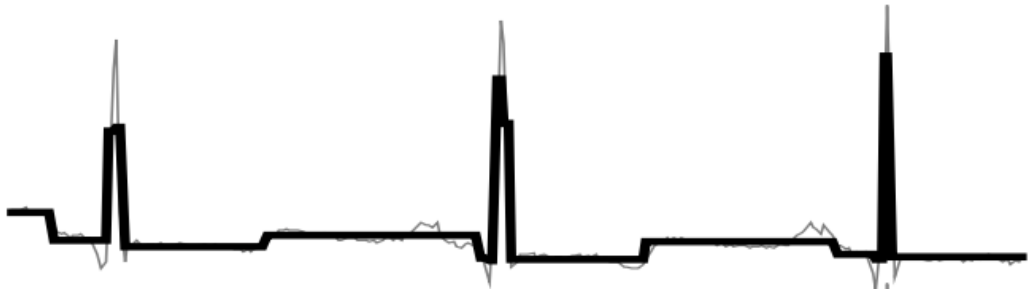
# Transformation-based representation




- Apply a set of *rules* to transform the whole sequence
- Non data-adaptive
  - Unchanged parameters
  - e.g., Piecewise Aggregate Approximation (PAA)
- Data-adaptive
  - Adaptive parameters
  - e.g., Adaptive Piecewise Constant Approximation (APCA)



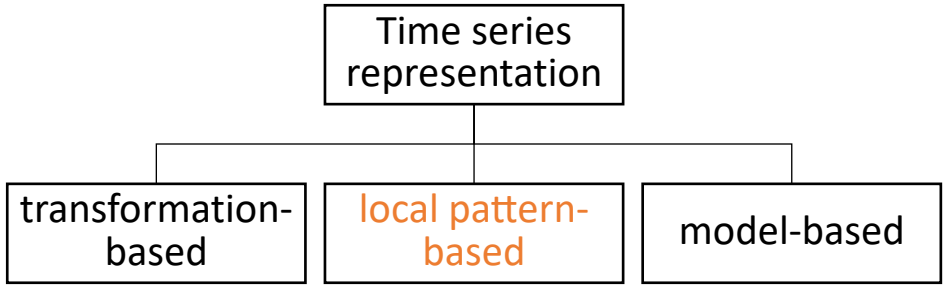
PAA (equal-length window)



APCA (adaptive-length window)

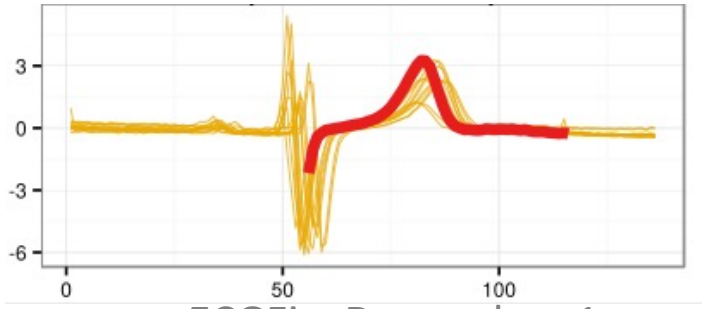
Piecewise transformations on ECG signal [Keogh et al., SIGMOD'01]

# Local pattern-based representation

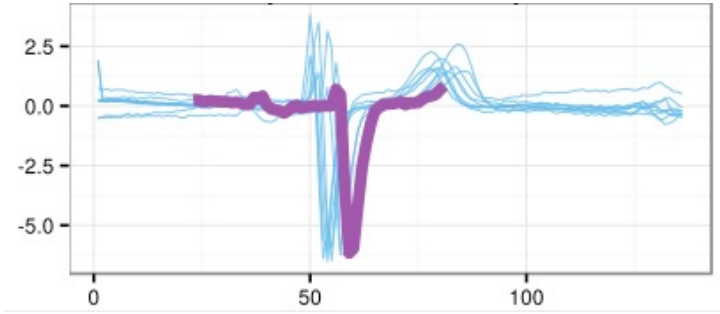



- Represent the whole sequence via (a set of) *local* patterns

- Discriminative patterns
  - Shapelets [Ye and Keogh, KDD'09]



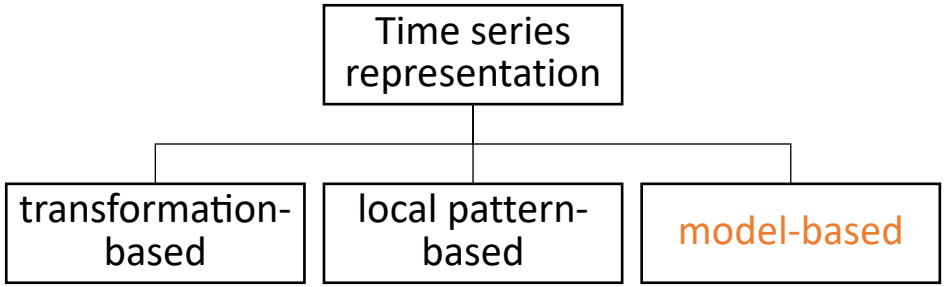

ECGFiveDays - class 1




ECGFiveDays - class 2

- Recurrent patterns
  - Frequent motifs [Wang et al., EDBT'16]

# Model-based representation




- Representing time series via model parameters

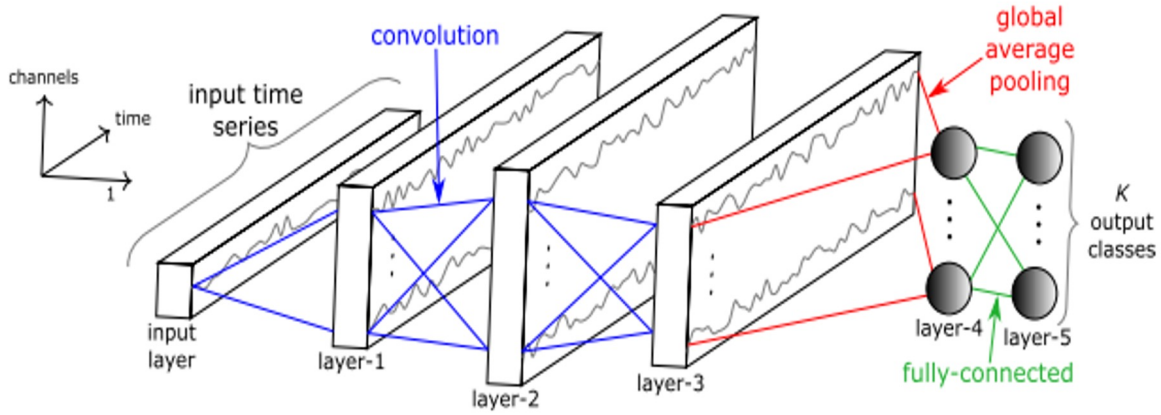
- Statistic modeling
  - e.g., Markov Chains (MCs)

$$P = (p_{ij}) = \begin{array}{c|cccc} & X_t & & & \\ X_{t-1} & 1 & 2 & \cdots & s \\ \hline 1 & p_{11} & p_{12} & \cdots & p_{1s} \\ 2 & p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & & & \cdots & \\ s & p_{s1} & p_{s2} & \cdots & p_{ss} \end{array}$$

where $p_{ij} = p(X_t = j \mid X_{t-1} = i)$

Representing TS as a transition probability matrix
[Sebastiani et al., IDA'99]

- Neural network-based modeling
  - Deep representations




Fully Convolutional Neural Network architecture
[Wang et al., IJCNN'17]

# Outline

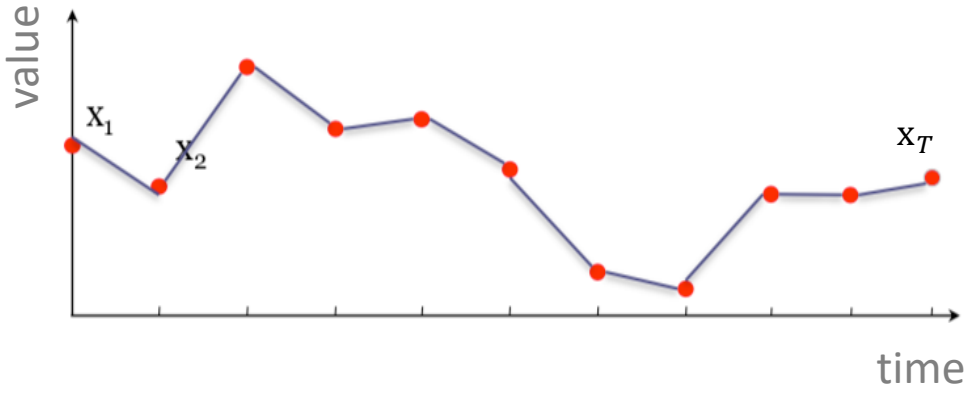
- Introduction
- Background
  - Time series mining
  - Time series representation
- ISMAP: Dynamic Feature Learning on Time Series Stream
- SMATE: Semi-supervised Learning on Multivariate Time Series
- GCN-M: Geo-located Time Series Forecasting with Missing Values
- Conclusion and perspectives

# Context & definitions

- Time series
  - Sequence of points ordered by time

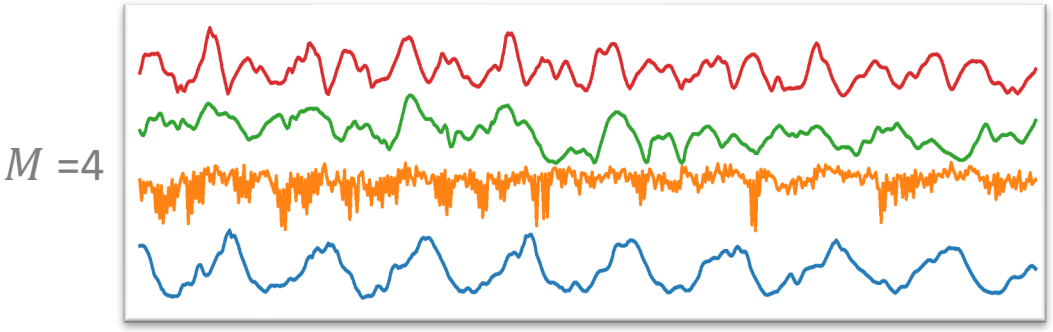
## Univariate Time Series (UTS)

$$x_1, \dots, x_T \in \mathcal{R}^M, M = 1$$




## Multivariate Time Series (MTS)

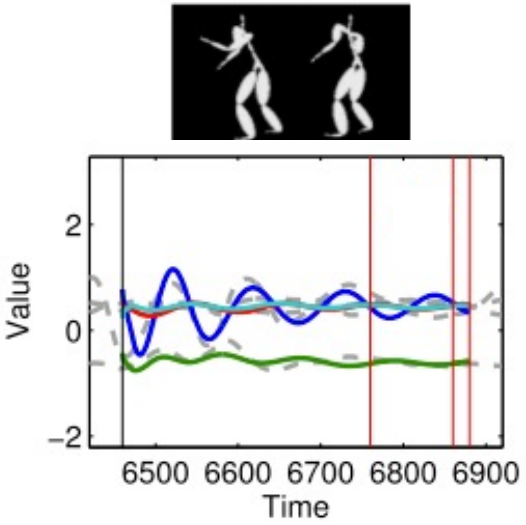
$$x_1, \dots, x_T \in \mathcal{R}^M, M > 1$$




Data from SHL-Huawei dataset

# Context & definitions

- Streaming context:
  - real-valued data flow (e.g., real-time sensor data)



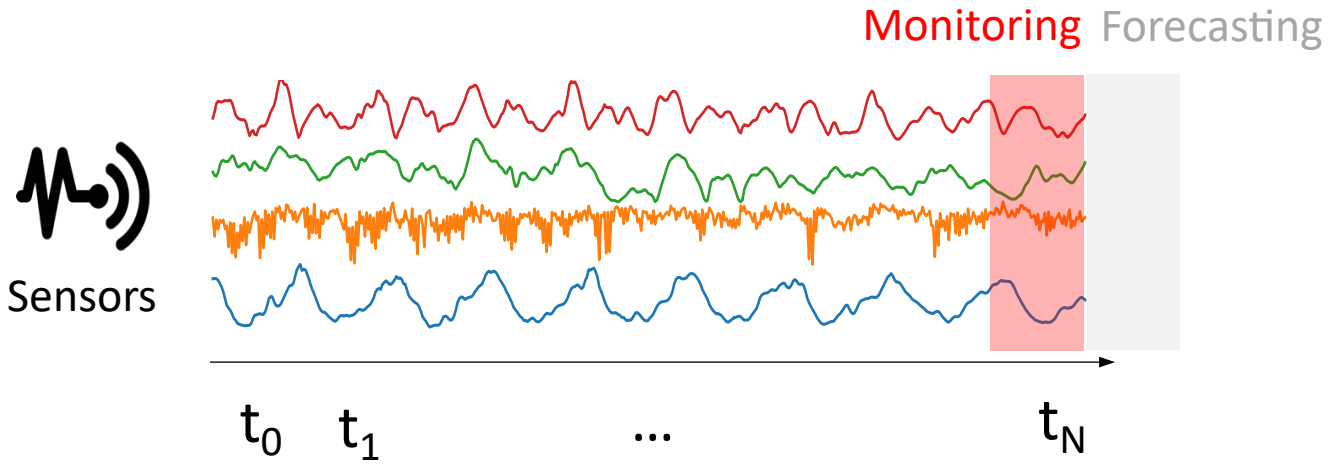

Real-time sensor data

- Time series in streaming context
  - Historical time series, i.e., offline time series
  - Streaming time series
  - Time series stream

# Context & definitions

- Streaming time series
  - A continuous input data stream where each instance is a real-valued data: $S=(t_1, t_2, \dots, t_N)$, where N is the time of the most recent input value.

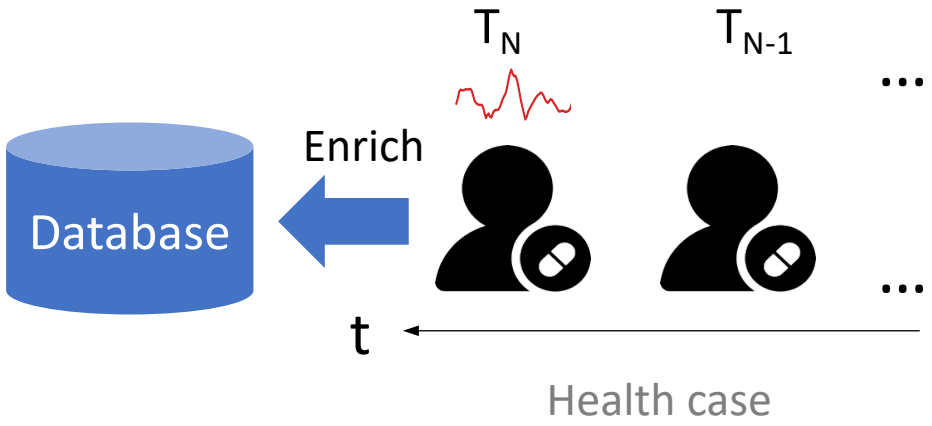
- Use cases:
  - Online monitoring
  - Real-time forecasting



# Context & definitions

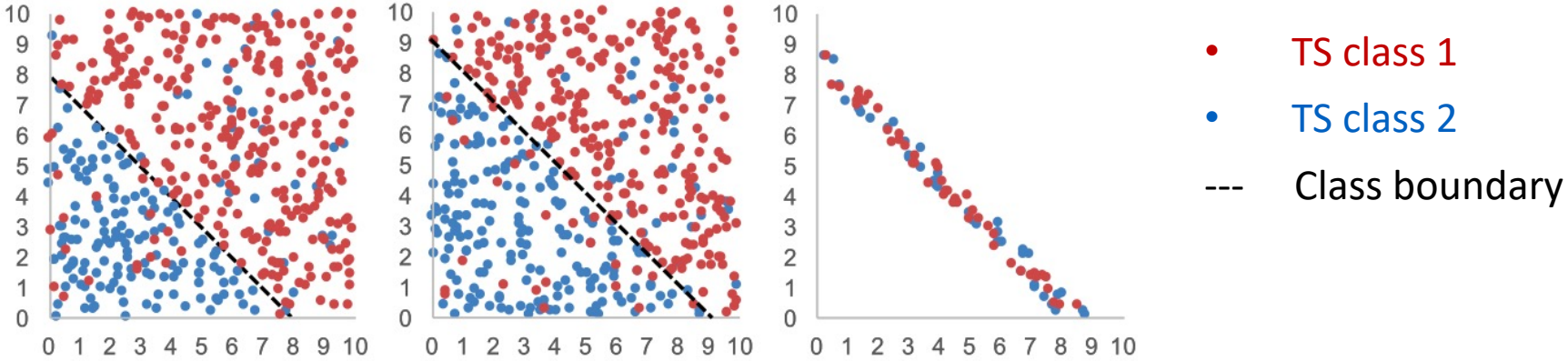
- Time series stream (our context)
  - A continuous input data stream where each instance is a time series: $S_{TS} = (T_1, T_2, \dots, T_N)$, notice that *N* increases with each new time-tick.

- Use cases:
  - Medical domain (e.g., ECG)
  - Astronomy discovery (e.g., Star Light Curves)



# Problem statement

- Complex temporal relationships in time series stream
  - Infinite length
  - Feature evolution
  - Concept drift



# Objectives

- TS features in streaming context
  - **Interpretability**: visually interpretable
  - **Incrementality**: feature extraction is incremental with new-coming instances [Feature Evolution]
  - **Adaptability**: adaptive to the evolving data distribution [Concept Drift]
- Learning model
  - **Scalability**
- Mainly designed for Time Series Classification (TSC) Task
  - **Training online**, classification on-line or off-line

# Related work

- Time series representation for classification

transformation-based

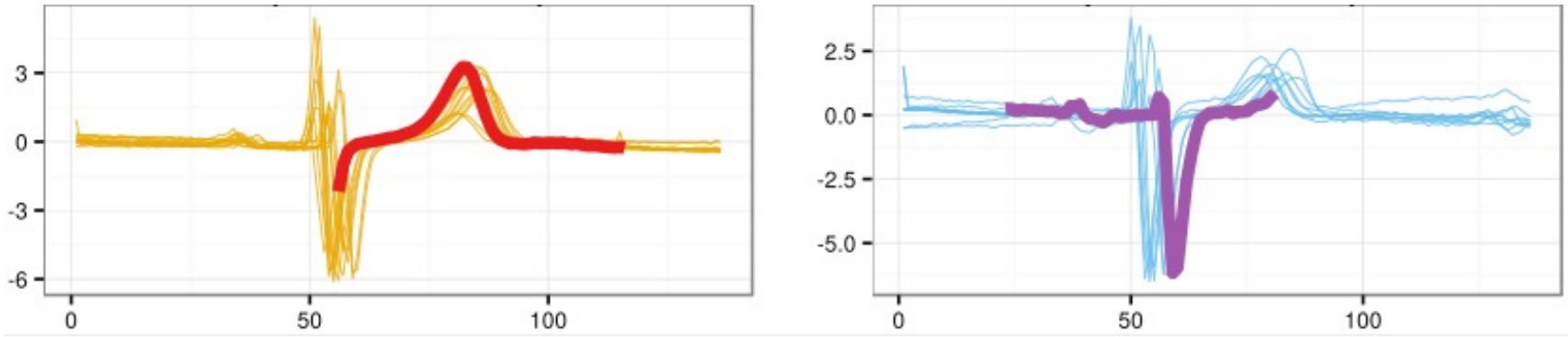
model-based

local pattern-based

| Feature representations | Classifier example | Related work |
|---|---|---|
| Raw representation | 1-NN | **1NN-ED,** **1NN-DTW** and its variants |
| Statistic summary | SVM or tree-based | **TSF** [Deng et al., Inf. Sci. 2013] |
| Deep representations | Neural Networks | **mWDN** [Wang et al., KDD'18], **InceptionTime** [Fawaz et al., DMKD'19], **LSTM-FCN** [Farim et al., arXiv'19] |
| Feature/model ensembles | Ensemble classifier | **BOSS** [Schäfer, DMKD'15] and its variants, **HIVE-COTE** [Lines et al., ICDM'17], **TDE** [Middlehurst et al., PKDD'20] |
| Local patterns | SVM or tree-based | **RPM** [Wang and Lin, EDBT'16], **Shapelet** [Ye and Keogh, KDD'09] and its variants |

# Why Shapelet[1] in our context?

- Definition
  - A representative shape in time series which is capable of distinguishing one class from the others

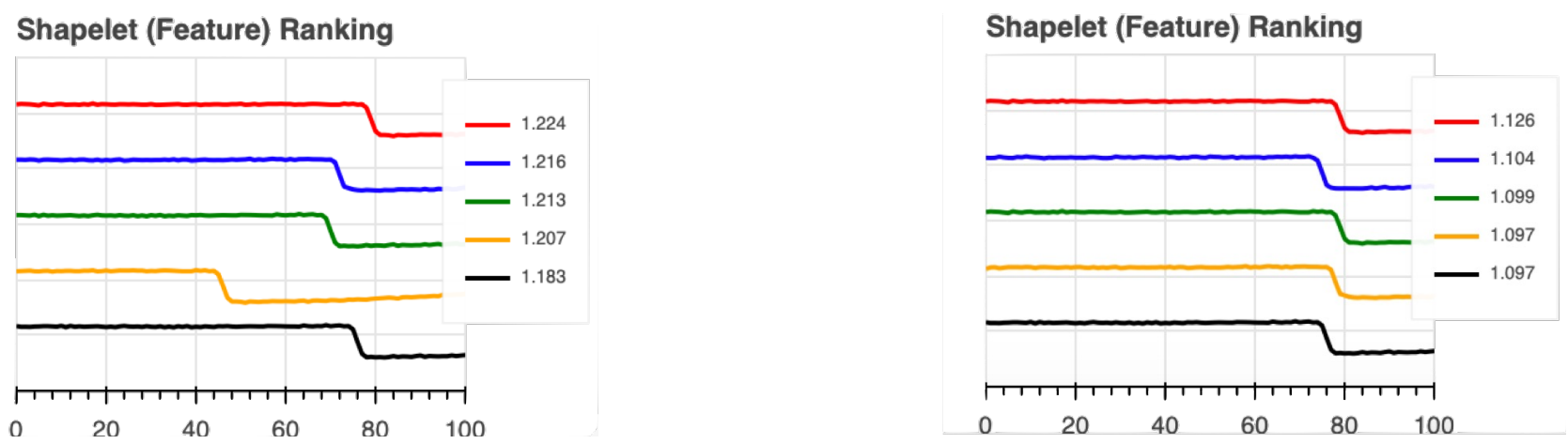



Most representative Shapelets in two classes from ECGFiveDays
[Wang and Lin, EDBT'16]

1. L. Ye and E. Keogh. "Time series shapelets: A New Primitive for Data Mining." In Proc. SIGKDD 2009

# Why Shapelet[1] in our context?

- Explainable for Feature Evolution in time series stream




$T_1 = 20$
Dataset *Trace*[2] (class 2)

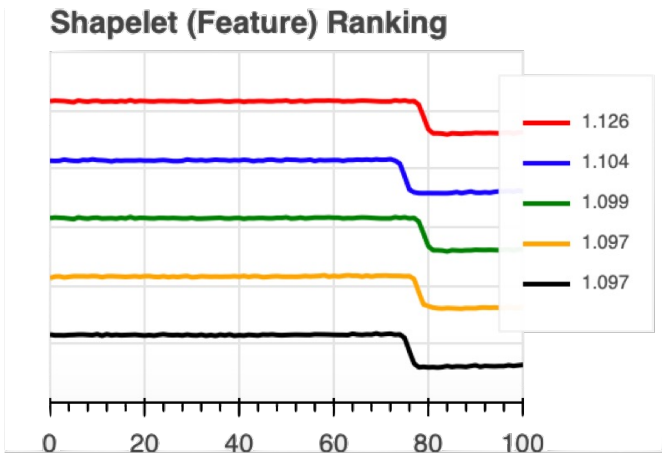
$T_2 = 100$
Dataset *Trace*[2] (class 2)

Feature Evolution
between $t \in [20, 100]$

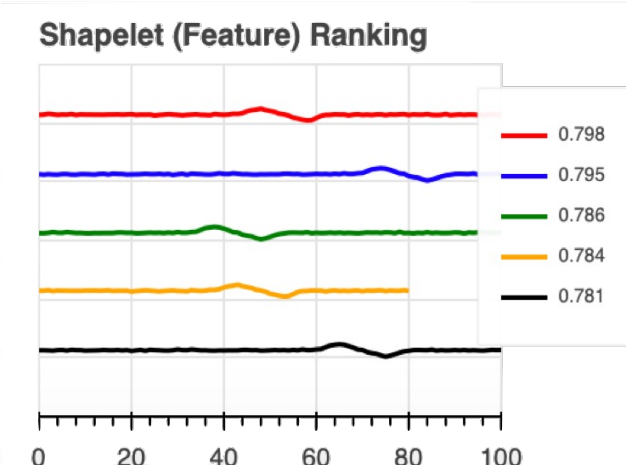
1. L. Ye and E. Keogh. "Time series shapelets: A New Primitive for Data Mining." In Proc. SIGKDD 2009
2. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

# Why Shapelet[1] in our context?

- Explainable for Concept Drift in time series stream




$T_2 = 100$
Dataset *Trace*[2] (class 2)

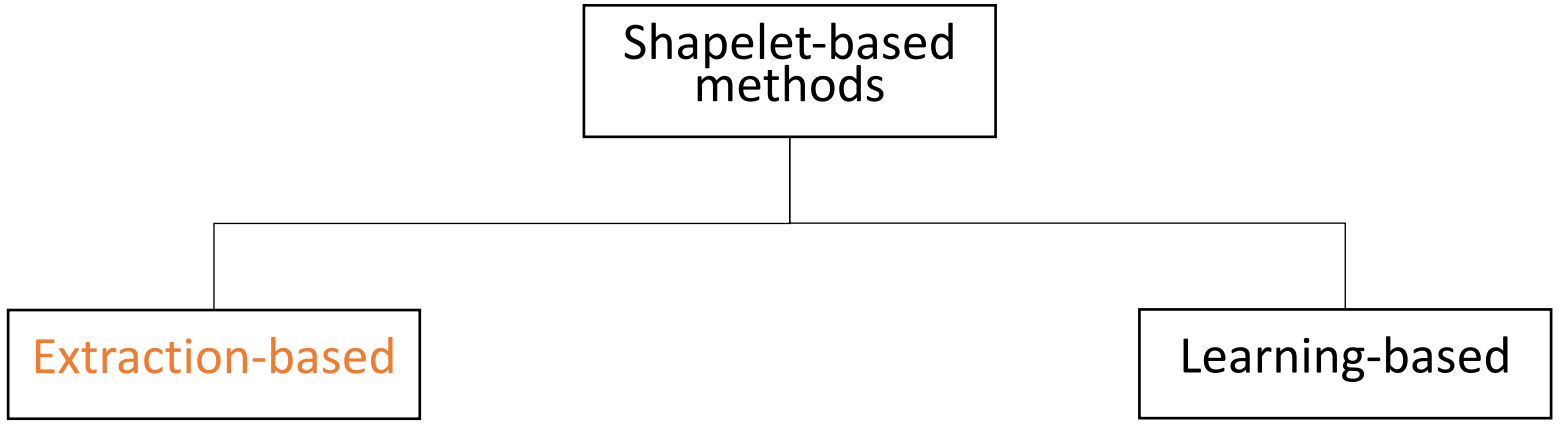



$T_3 = 200$
Dataset *Trace*[2] (class 2)

Concept Drift between
$t \in [100, 200]$

1. L. Ye and E. Keogh. "Time series shapelets: A New Primitive for Data Mining." In Proc. SIGKDD 2009
2. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

# Shapelet-based methods




- Highly interpretable (decision-tree)

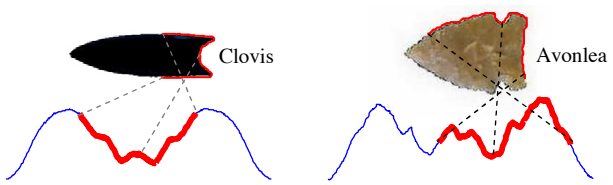



Figure from [Ye and Keogh, KDD'09]

- End-to-end (gradient-based learning)
- Generally not interpretable

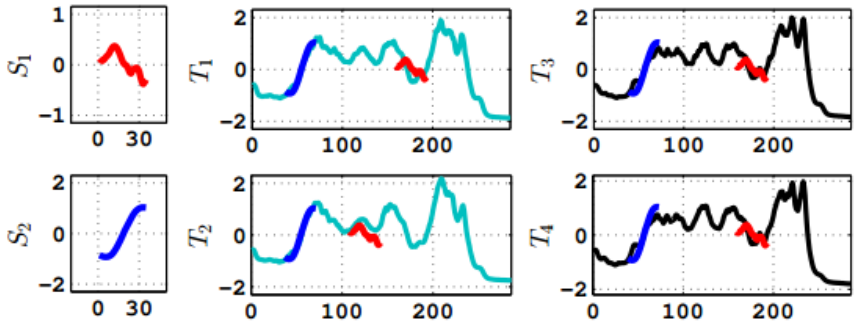



Figure from [Grabocka et al., KDD'14]

# Algorithm for Shapelet Extraction

- Distance Profile & Matrix Profile[1]

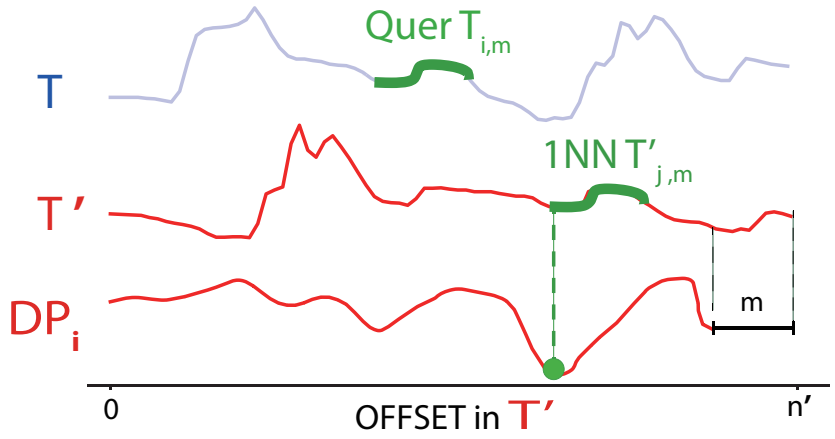



**Figure 2.1:** ***Distance Profile*** *between Query $T_{i,m}$ and target time series $T'$, where $n'$ is the length of $T'$. $DP_{i,j}$ can be considered as a meta TS annotating target $T'$*

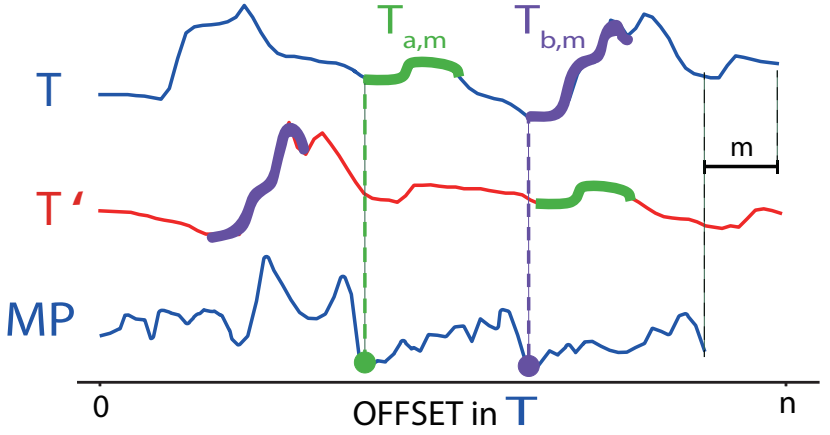



**Figure 2.2:** ***Matrix Profile*** *between Source $T$ and Target $T'$, where $n$ is the length of $T$. Intuitively, $MP_i$ shares the same offset as source $T$*

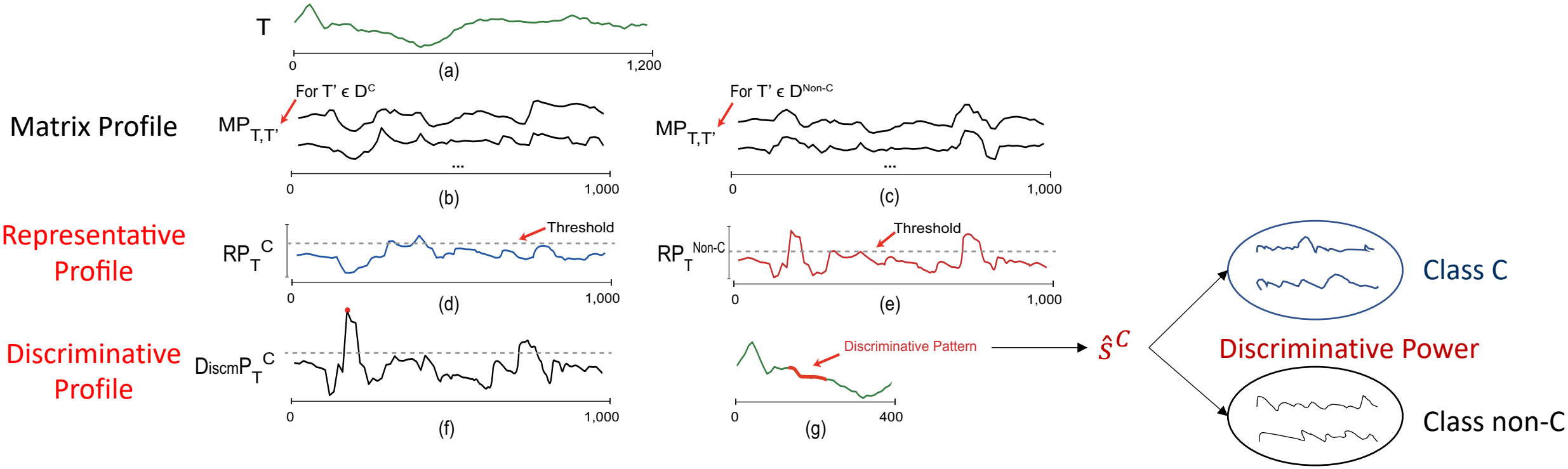
➢ Find the Nearest Neighbor of the Query

➢ Find the closest pairs between two TS

1. Chin-Chia Michael Yeh et al. "Matrix Pro le I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets." In Proc. ICDM 2016

# Proposal - SMAP

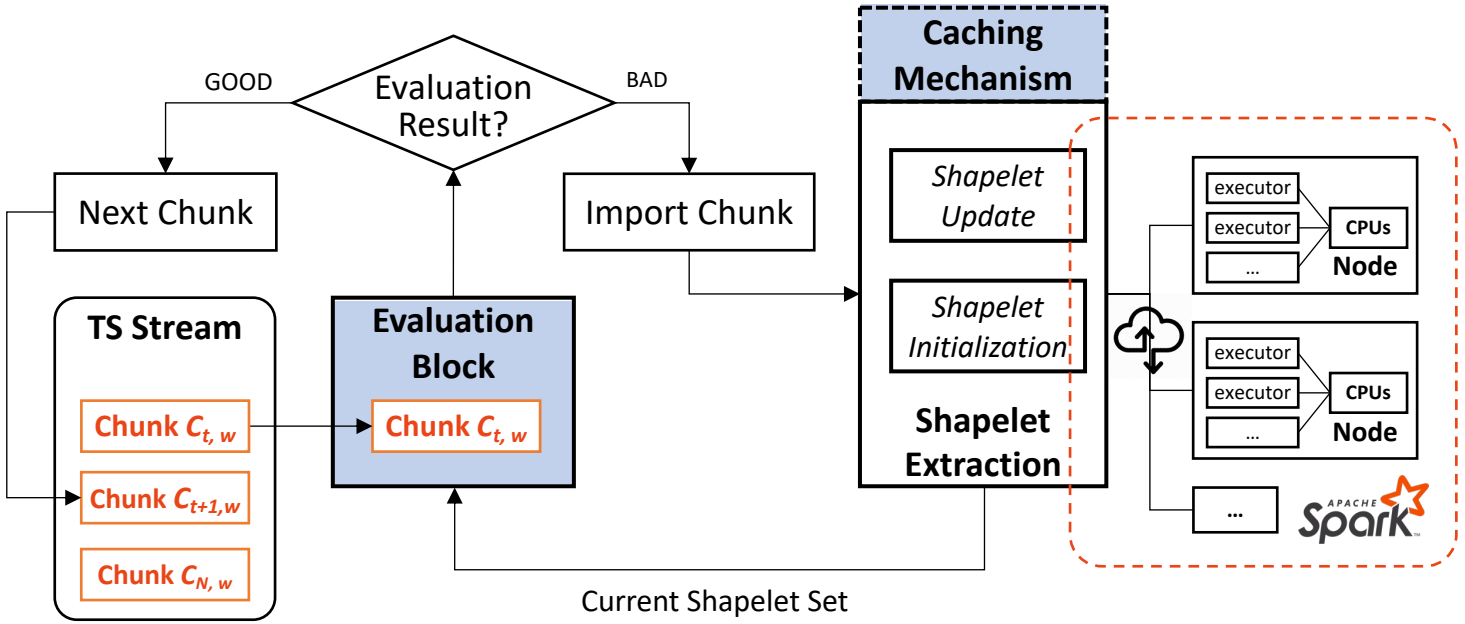
- SMAP[1] : Shapelet Extraction on Matrix Profile




1. J. Zuo, K. Zeitouni and Y. Taher, Exploring interpretable features for large time series with SE4TeC. In Proc. EDBT 2019

# Proposal - Incremental version of SMAP

- ISMAP[1] : **Incremental** and **adaptive** Shapelet Extraction on Matrix Profile



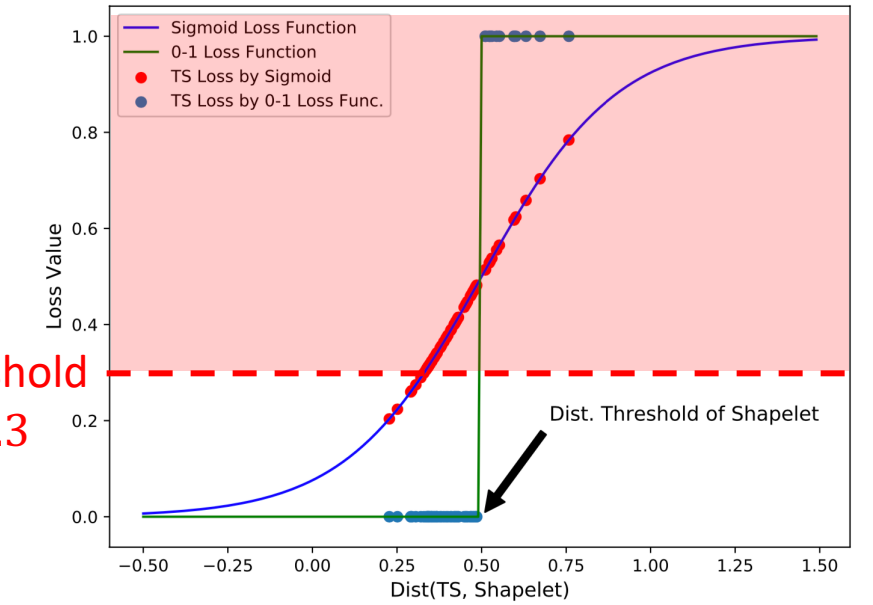

Test-then-Train strategy

1. J. Zuo, K. Zeitouni and Y. Taher, Incremental and Adaptive Feature Exploration over Time Series Stream, IEEE Big Data 2019

# ISMAP - Evaluation Block

## Shapelet Evaluation

e.g., loss threshold $\Delta = 0.3$




Shapelet Evaluation over newly input TS instances

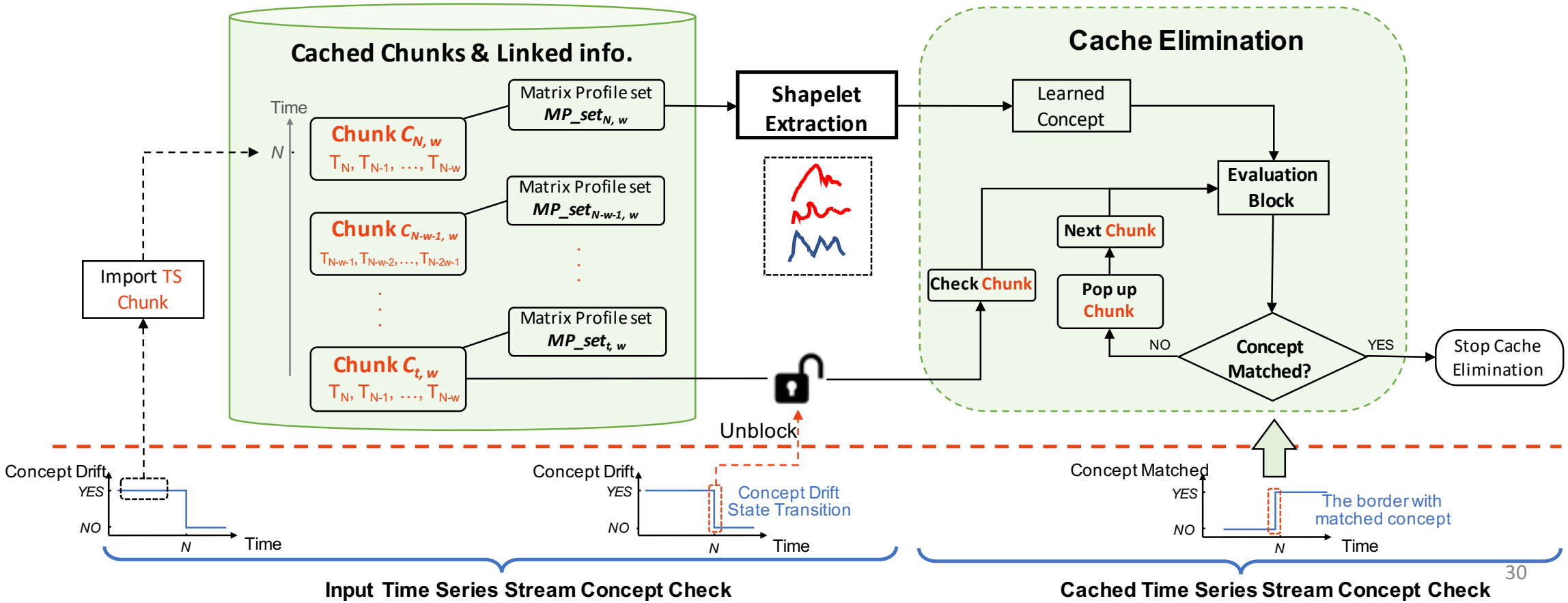
## Concept Drift detection

Consider the evaluation loss as a signal

- Page-Hinkey (PH) Test: initially designed for change point detection in signal processing.
- $\lambda$: *PH threshold* to detect a Concept Drift
- $$Concept\ Drift = \begin{cases} True, & PH_N \geq \lambda \\ False, & otherwise \end{cases}$$

# ISMAP - Elastic Caching Mechanism

- One-pass algorithm
  - Only conserve the data under the current concept to be learned
- Conserve the historical Shapelets in the out-of-date concepts (optional)



# Experiments

## Research Questions:

- RQ1. Incremental learning with ISMAP
  - Stable-concept time series stream
  - To validate the incremental behavior

→

Datasets:
- 14 datasets from UCR Archive[1]

Baselines (Shapelet Tree classifiers):
- Information Gain (IG) [Ye and Keogh, KDD'09]
- Kruskall-Wallis (KW), Mood's Median (MM) [Lines and Bagnall, IDEAL'12]

- RQ2. Adaptive learning with ISMAP
  - Drifting-concept time series stream
  - To validate the drift detection behavior and elastic caching mechanism

→

Datasets:
- Synthetic *Trace* and *ECG5000* datasets[1]:
  - Randomly put noise for Data Augmentation
  - Two concept drifts are inserted in each dataset

1. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

# RQ1. Incremental learning with ISMAP

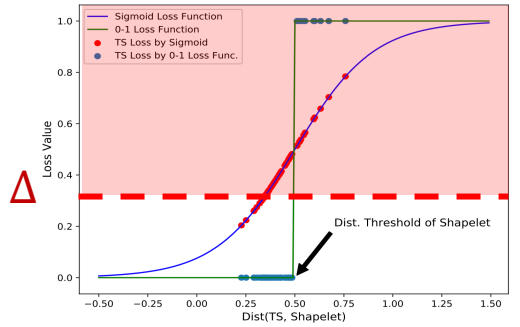
- Incremental behavior
  - Captured by $Compression\ Ratio = \frac{N_{import}}{N_{training}}$
- Possible to combine with other TS classifiers:
  - Shapelet Transform [Lines et al., KDD'12]
  - HIVE-COTE [Lines et al., ICDM'16]

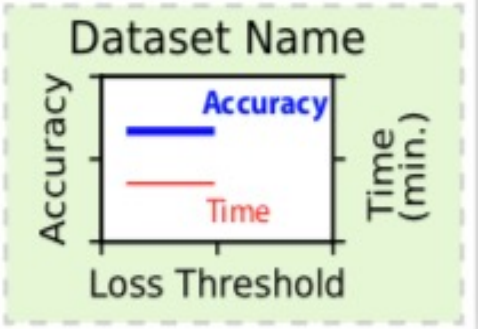
| Type | Name | Train/Test | Class | Length | IG | KW | MM | ISMAP(best) | Para. ($\Delta$) | Comp. Ratio |
|---|---|---|---|---|---|---|---|---|---|---|
| Simulated | SyntheticControl | 300/300 | 6 | 60 | **0.9433** | 0.9000 | 0.8133 | 0.7007 | 0.35 | 46.7% |
| Sensor | Trace | 100/100 | 4 | 275 | 0.9800 | 0.9400 | 0.9200 | **1** | 0.5, 0.45 | 26.0% |
| | MoteStrain | 20/1252 | 2 | 84 | 0.8251 | 0.8395 | 0.8395 | **0.9169** | 0.45 | 60.0% |
| | SonyAIBO.I | 20/601 | 2 | 70 | 0.8453 | 0.7281 | 0.7521 | **0.9151** | 0.4 | 95.0% |
| | SonyAIBO.II | 27/953 | 2 | 65 | 0.8457 | - | - | **0.8583** | 0.4 | 63.0% |
| | ItalyPower. | 67/1029 | 2 | 24 | 0.8921 | 0.9096 | 0.8678 | **0.9466** | 0.45 | 25.4% |
| ECG | ECG5000 | 500/4500 | 5 | 140 | 0.7852 | - | - | **0.9109** | 0.4 | 9.4% |
| | ECGFiveDays | 23/861 | 2 | 136 | 0.7747 | 0.8721 | 0.8432 | **0.9826** | 0.4 | 51.2% |
| | TwoLeadECG | 23/1189 | 2 | 82 | 0.8507 | 0.7538 | 7657 | **0.9337** | 0.5 | 47.8% |
| Images | Symbols | 25/995 | 6 | 398 | 0.7799 | 0.5568 | 0.5799 | **0.8113** | 0.35 | 96.0% |
| | Coffee | 28/28 | 2 | 286 | **0.9643** | 0.8571 | 0.8671 | 0.9286 | 0.4 | 78.6% |
| | FaceFour | 24/88 | 4 | 350 | 0.8409 | 0.4432 | 0.4205 | **0.9886** | except 0.45 | 62.5% |
| | DiatomSize. | 16/306 | 4 | 345 | 0.7222 | 0.6111 | 0.4608 | **0.8758** | 0.5 | 50.0% |
| Motion | GunPoint | 50/150 | 2 | 150 | 0.8933 | 0.9400 | 0.9000 | **0.9733** | 0.45 | 42.0% |

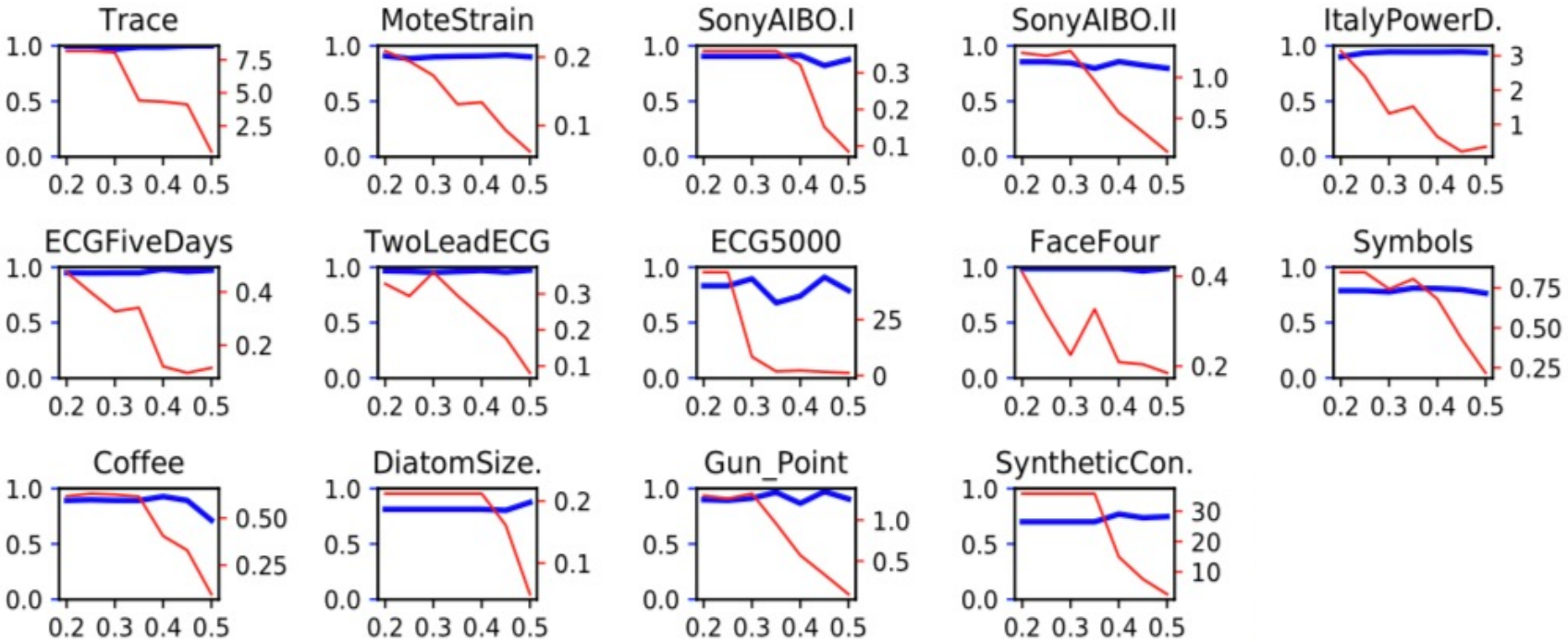
1. J. Lines, L. M. Davis, J. Hills, and A. Bagnall, "A shapelet transform for time series classification," in Proc. SIGKDD 2012
2. J. Lines, S. Taylor, and A. Bagnall, "Hive-cote: The hierarchical vote collective of transformation-based ensembles for time series classification," IEEE ICDM 2016

# RQ1. Incremental learning with ISMAP

- Trade-off between Accuracy and Loss Threshold Δ




## In theory

- Loss threshold ↗, the efficiency ↗, the accuracy ↘

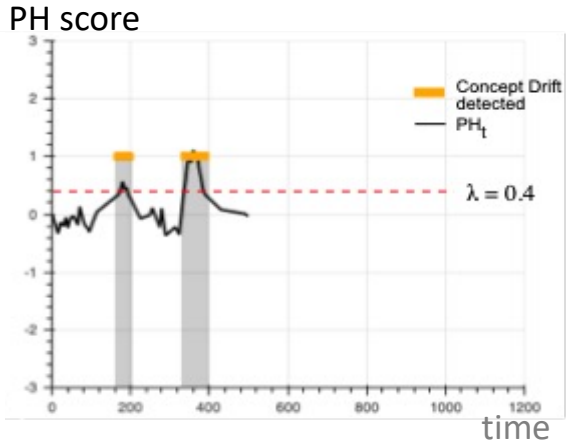
## In practice

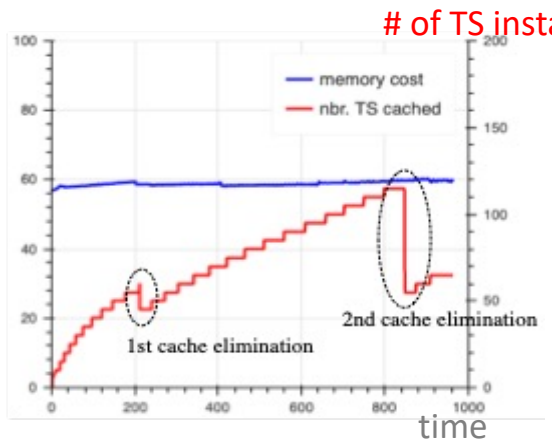
- The highest accuracy falls in the range $\Delta \in [0.35, 0.45]$.

# RQ2. Adaptive learning with ISMAP

- Concept drift detection & Elastic caching mechanism[1]

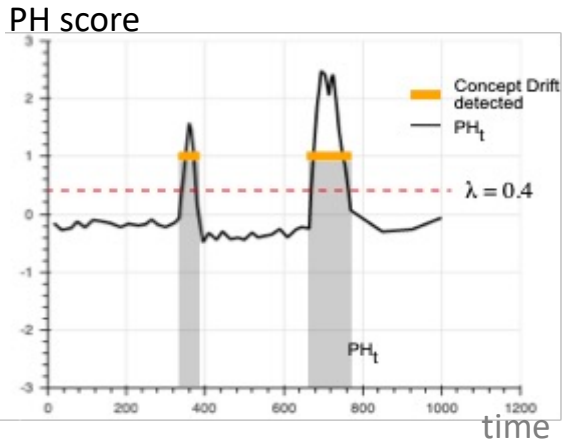
ECG5000

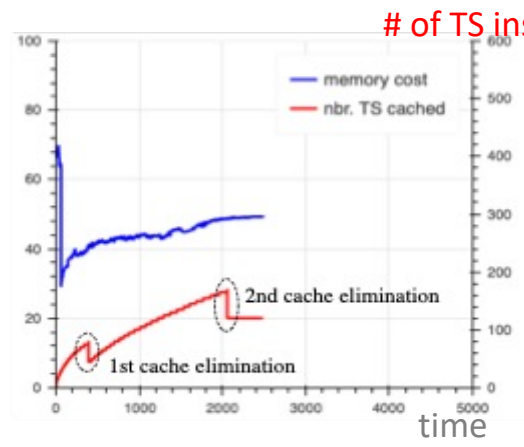



- Two concept drifts detected
- 65 of 500 instances cached

Trace




- Two concept drifts detected
- 120 of 1000 instances cached

1. J. Zuo, K. Zeitouni, and Y. Taher, “ISETS: Incremental Shapelet Extraction from Time Series Stream”, demo paper in ECML-PKDD’19

# ISMAP - Conclusion

- Shapelet representation is natively interpretable for explaining the feature evolution and concept drift in the time series stream.
- Our proposal *ISMAP* extracts incremental and adaptive Shapelets from the time series stream
- Our proposed *elastic caching mechanism* handles the infinite time series stream.
- ISMAP is applicable in the scenarios where:
  - New TS instances enrich the learned concept
  - New TS instances may lead to Concept Drift

Github page

# Outline

- Introduction
- Background
  - Time series mining
  - Time series representation
- ISMAP: Dynamic Feature Learning on Time Series Stream
- SMATE: Semi-supervised Learning on Multivariate Time Series
- GCN-M: Geo-located Time Series Forecasting with Missing Values
- Conclusion and perspectives

# Problem statement & objectives

- Complex structure in multivariate time series (MTS)
  - Temporal relationships
  - Inter-variable relationships
- Costly labeling on multivariate time series

## Objectives

- Learn an appropriate MTS representation
  - The **temporal dependencies**: temporal dynamics
  - The **dynamic interaction**s between the variables: spatial dynamics
- Semi-supervised representation learning
  - Explore thoroughly the information in labeled and unlabeled samples

# Related work – MTS representations

1. Combine features from each variable (i.e., 1-D series)
   - Global features: **1NN-DTW$_I$** [Yekta et al., DMKD'15]
   - Shapelet features: **Shapelet Ensemble** [Cetin et al., SDM'15], **M-Shapelet Discovery** [Grabocka et al., KAIS'16]
   - Motif features: **WEASEL+MUSE** [Schafer et al., AALTD'18], **Global Discriminative Patterns** [Nayak et al., SDM'18]
   - Deep Representation features: **Multi-Channels CNN** [Zheng et al., WAIM'14]
2. Extract features directly from all variables
   - Global features: **1NN-DTW$_D$** (**1NN-DTW$_A$**) [Yekta et al., DMKD'15]
   - Motif features: Symbolic Representation for MTS (**SMTS**) [Baydogan et al., DMKD'15]
   - Deep Representation features: Modified DNN approaches for Univariate TSC, e.g., **LSTM-FCNs** [Karim et al., ArXiv'19], **InceptionTime** [Fawaz et al., ArXiv'19], **ROCKET** [Dempster et al., ArXiv'19], etc.
3. Consider the interactions between the variables
   - Variable correlation: **MLSTM-FCNs** [Karim et al., Neural Networks'19]
   - Attention Mechanism: **CA-SFCN** [Hao et al. IJCAI'20]
   - 2D-CNN with 1D-CNN: **MTEX-CNN** [Assaf et al., ICDM'19], **XCM** [Fauvel et al., ArXiv'20]
   - Graph Pooling: **MTPool** [Xu et al., ArXiv'20]

All these approaches are fully supervised

# Related work – Semi-supervised learning on TS

## Semi-supervised Learning on UTS

- Self-training or Positive Unlabeled Learning
  - **SSTSC** [Wei and Keogh, KDD'06]
  - **LCLC** [Nguyen et al., IJCAI'11]
  - **DTW-D** [Chen et al., KDD'13], etc.
  - **SSSL** [Wang et al., PR'19]
- Clustering based
  - **SUCCESS** [Marussy et al., ICAISC'13]
- Self-Supervised Learning
  - **MTL** [Javed et al., PAKDD'20]

Most of the semi-supervised approaches are applied in Univariate Time Series

## Semi-supervised Learning on MTS

- **USRL** [Franceschi et al., NeurIPS'19]: contrastive learning
- **TapNet** [Zhang et al., AAAI'20]: attentional prototypical network

# Proposal - SMATE

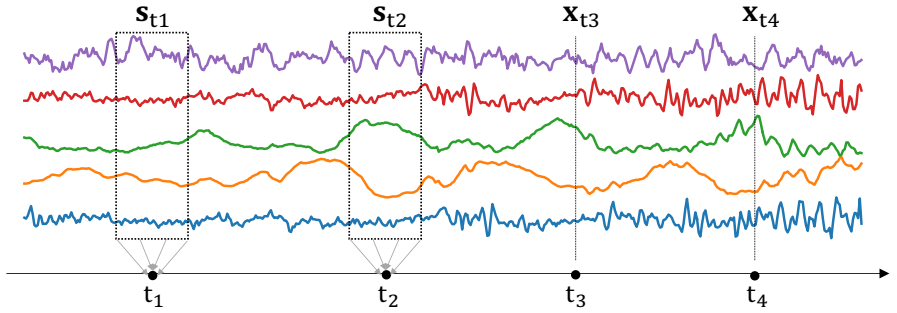
- Semi-supervised Spatio-Temporal Representation Learning on Multivariate Time Series[1]
  - **Representation Learning** on $\mathbf{x} \in \mathbb{R}^{T \times M}$
    - Learn a low-dimensional representation $\mathbf{h} \in \mathbb{R}^{L \times D}$, where $L < T,\ D < M$
    - $\mathbf{h}$ embeds the spatial and temporal features of $\mathbf{x}$
  - **Semi-supervised regularization** in the embedding space $\mathcal{H}$
    - Combine both labelled and unlabelled samples
    - Learn class-separable representations for downstream tasks, e.g., MTS classification

1. J. Zuo, K. Zeitouni, Y. Taher, SMATE: Semi-supervised Spatio-Temporal Representation Learning on Multivariate Time Series, IEEE ICDM 2021

# Spatio-Temporal Representation on MTS

## Intuition

- System status at time t
  - Local value $\mathrm{x}_t \in \mathbb{R}^M$
  - Neighbor values $s_t = [x_{t-m/2}, x_{t+m/2}]$
- Spatio-temporal features
  - Temporal Dynamic $p(\mathrm{x}_{t'} \mid \mathrm{x}_t)$
  - Spatial Dynamic $p(\mathrm{s}_{t'} \mid \mathrm{s}_t)$
  - Spatio-temporal dynamics $(\mathrm{x}_t, \mathrm{s}_t) \rightarrow (\mathrm{x}_{t'}, \mathrm{s}_{t'})$

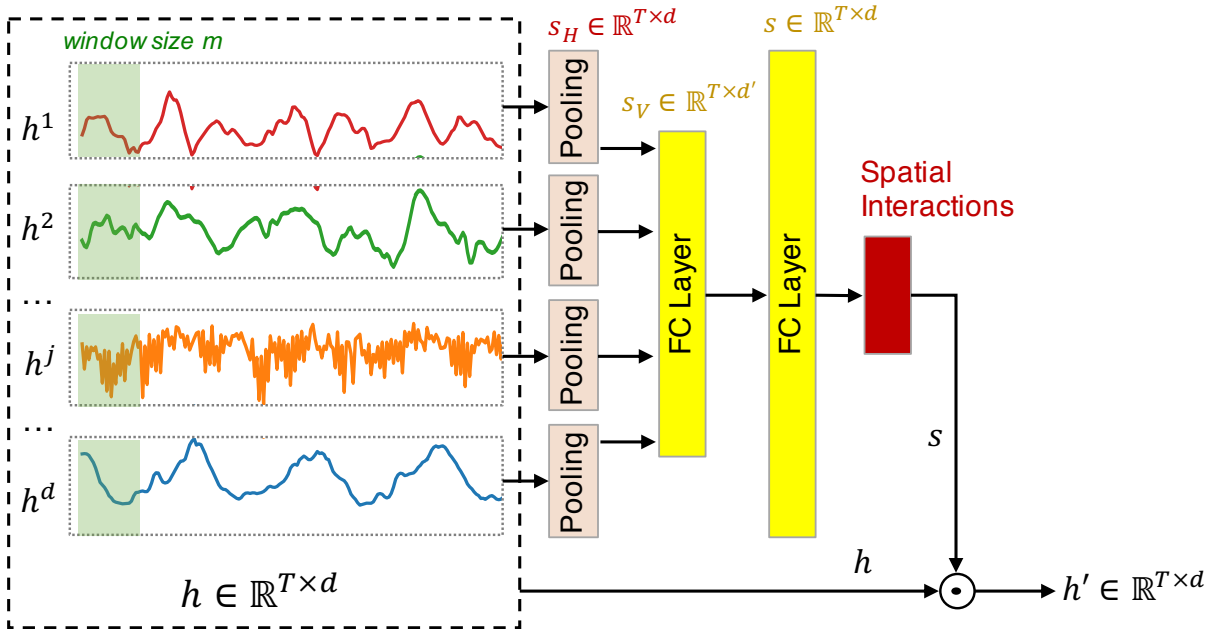



The spatial and temporal structure in a MTS sample representing "Walking" activity (with 5 sensors) of the SHL[1] dataset.

1. H. Gjoreski, M. Ciliberto, L. Wang, F. J. O. Morales, S. Mekki, S. Valentin, D. Roggen. "The University of Sussex-Huawei Locomotion and Transportation Dataset for Multimodal Analytics with Mobile Devices." IEEE Access 6 (2018): 42592-4260

# Spatio-Temporal Representation on MTS

- Spatial Modelling Block (SMB)
  - Capture the spatial interaction at **segment level**
  - 1-d average Pooling: output the **temporal status**
  - 2 Fully Connected (FC) layers: **interacting** the temporal status **in spatial direction**
  - **Output:** the weighted MTS considering the spatial interaction

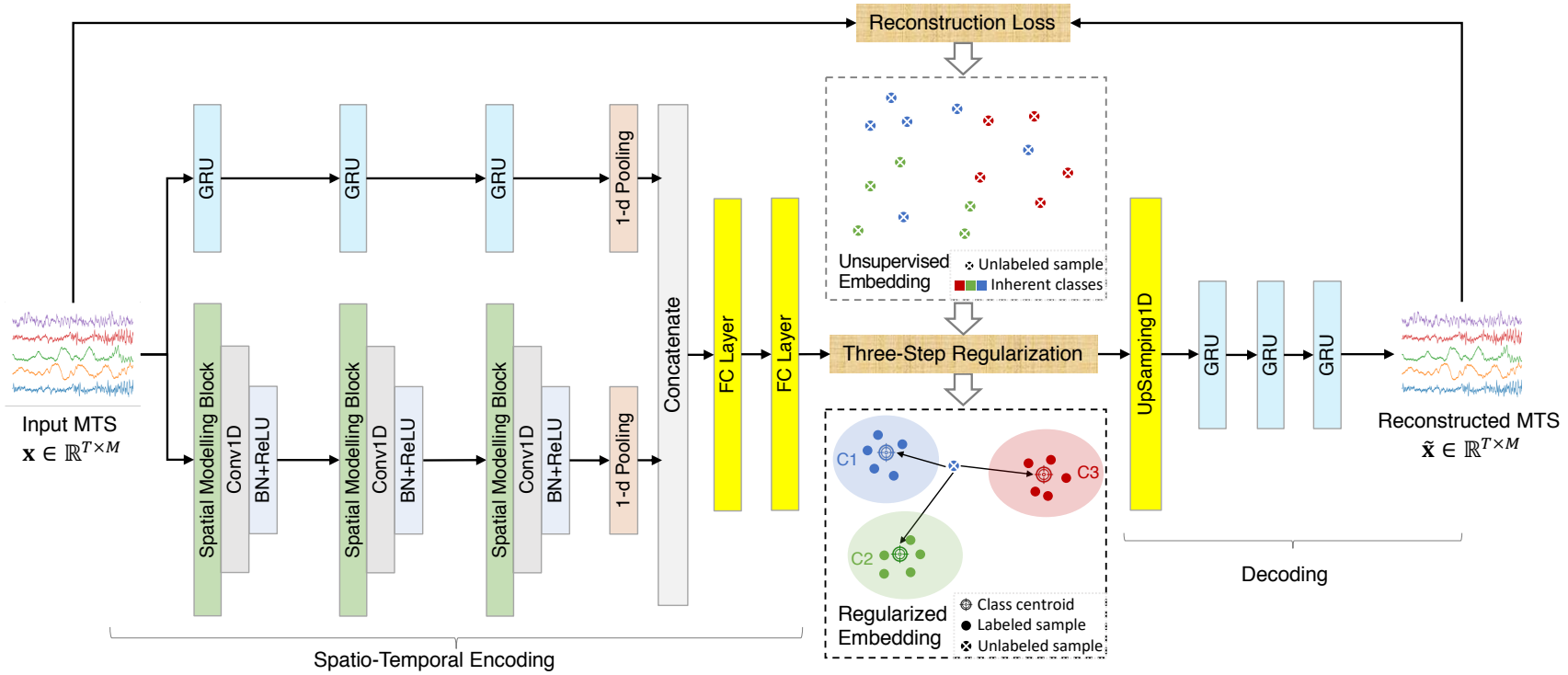



$s_{H_i} \in \mathbb{R}^d$
- Horizontal temporal status

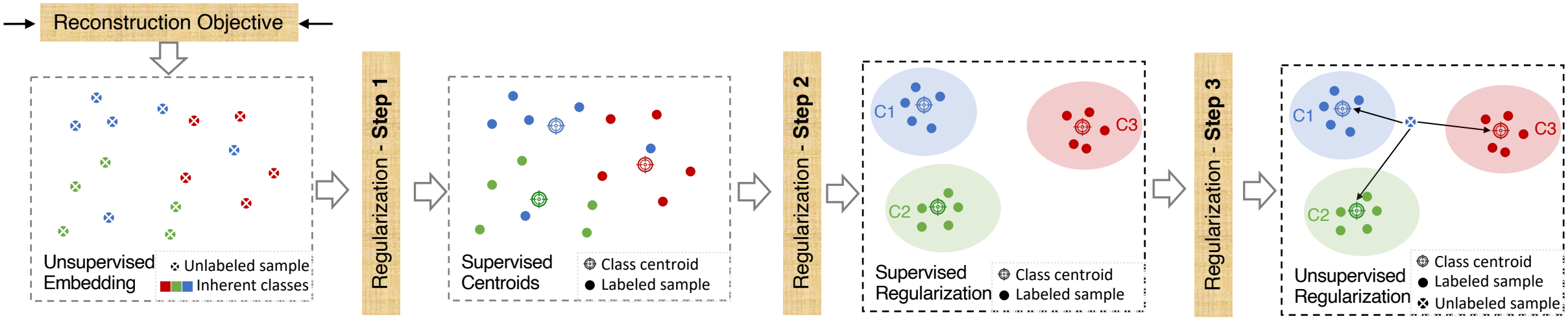
$s_i \in \mathbb{R}^d$
- Spatial interaction weights

# SMATE - Model Structure

- Based on an **asymmetric auto-encoder** structure
- Two channels for **Spatio-temporal encoding**
  - GRU: $p(\mathrm{x}_{t'} \mid \mathrm{x}_t)$
  - SMB + Conv1D: $p(\mathrm{s}_{t'} \mid \mathrm{s}_t)$
- **Three-Step Regularization**



# SMATE - Three-Step Regularization




## Step 1: Supervised Centroids Initialization

The embedding collection of $X^k$ :

- $H^k \in \mathbb{R}^{N_k \times L \times D}$

The centroid of class *k*

- $c_k = mean(H^k)$, $c^k \in \mathbb{R}^{L \times D}$

## Step 2: Supervised Centroids Adjustment

**Intuition:**

- The *near-by* samples have *larger contribution weights* to the class centroids

$$W_{k,i} = 1 - \frac{dist(h_\theta(\mathbf{x}_i), c_k)}{\sum_{j=1}^{K} dist(h_\theta(\mathbf{x}_i), c_j)}$$

$$c_k = \sum_{i=1}^{N_k} W_{k,i} \cdot h_i^k, \quad h_i^k \in H^k$$

## Step 3: Unsupervised Centroids Adjustment

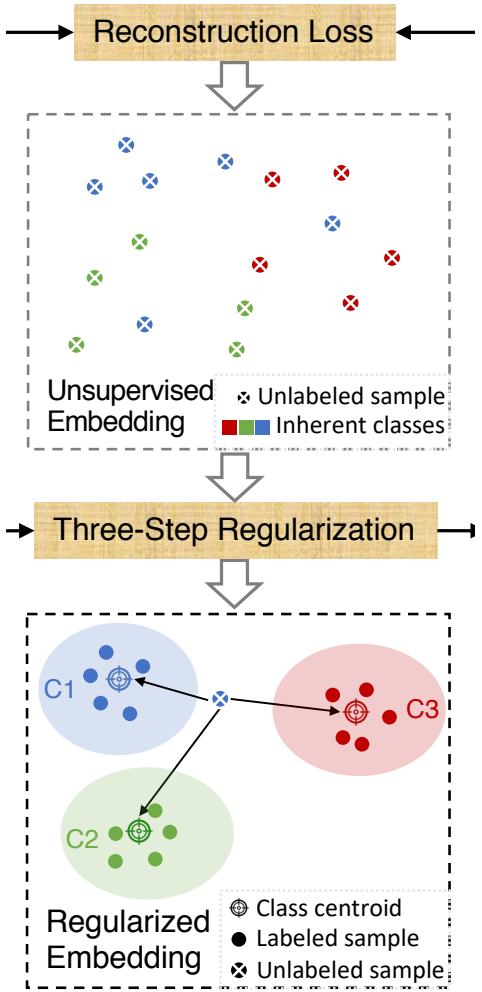
- The propagated label from the distance-based class probability

$$\hat{p}_\theta(y = k|\hat{\mathbf{x}}_i) = 1 - \frac{dist(h_\theta(\hat{\mathbf{x}}_i), c_k)}{\sum_{j=1}^{K} dist(h_\theta(\hat{\mathbf{x}}_i), c_j)}$$

- The class centroid $c_k$ is further adjusted:

$$c_k = \frac{N_k}{N_k + \hat{N}_k} \sum_{i=1}^{N_k} W_{k,i} \cdot h_i^k + \frac{\hat{N}_k}{N_k + \hat{N}_k} \sum_{i=1}^{\hat{N}_k} \hat{p}_{k,i} \cdot \hat{h}_i^k$$

# SMATE - Joint Model Optimization




## Regularization loss:

- With labelled samples
- With class centroids regularized by both labelled and unlabelled samples

$$L_{Reg}(\theta) = -\sum_{k} logW_{\theta}(y = k|\mathbf{x})$$

## Reconstruction loss:

$$L_R = \mathbb{E}_{\mathbf{x}_{1:T}}\left[\sum_{t} \|\mathbf{x}_t - \tilde{\mathbf{x}}_t\|_2\right]$$

## Objective function:

$$\min_{\theta}(L_R + \lambda L_{Reg})$$

# Experiments

Research Questions:

- RQ1. Classification performance
- RQ2. Semi-supervised classification performance
- RQ3. Model efficiency
- RQ4. Interpretation over the representation space
- RQ5. Performance of the Spatial Modeling Block

# RQ1. Classification performance

- Evaluating SMATE for fully supervised representation learning

**30 datasets from UEA Archive[1]**

**13 baselines:**

- Distance-based 1-NN classifier on non-normalized (*non-norm*) or normalized (*norm*) MTS
  - **1NN-ED** (*non-norm & norm*)
  - **1NN-DTW$_I$** (*non-norm & norm*); **1NN-DTW$_D$** (*non-norm & norm*)
  - **1NN-DTW$_A$** (*norm*) [Yekta et al., DMKD'15]
- Bag-of-patterns classifier
  - **WEASEL+MUSE** [Schäfer et al., AALTD'18]
- Deep Learning-based classifier:
  - **SMATE$_{NR}$**: SMATE without supervised Regularization
  - **MLSTM-FCNs** [Karim et al., Neural Networks'19]**, USRL**[Franceschi et al., NeurIPS'19]**, TapNet** [Zhang et al., AAAI'20], **CA-SFCN** [Hao et al., IJCAI'20]

1. www.timeseriesclassification.com

# RQ1. Classification performance

- Fully supervised representation learning
  - SVM on the learned representation

SMATE performs the best among all the baselines, especially on EEG/MEG data

Performance Comparison for MTS classification over UEA MTS archive

| Dataset | SMATE | $SMATE_{NR}$ | USRL | TapNet | MLSTM -FCN | CA-SFCN | WEASEL +MUSE | 1NN-ED | 1NN-$DTW_I$ | 1NN-$DTW_D$ | 1NN-ED (*norm*) | 1NN-$DTW_I$ (*norm*) | 1NN-$DTW_D$ (*norm*) | 1NN-$DTW_A$ (*norm*) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ArticularyWordR. | **0.993** | 0.987 | 0.987 | 0.987 | 0.973 | 0.97 | 0.99 | 0.97 | 0.98 | 0.987 | 0.97 | 0.98 | 0.987 | 0.987 |
| AtrialFibrillation | 0.133 | 0.133 | 0.133 | **0.333** | 0.267 | **0.333** | **0.333** | 0.267 | 0.267 | 0.2 | 0.267 | 0.267 | 0.22 | 0.267 |
| BasicMotions | **1** | **1** | **1** | **1** | 0.95 | **1** | **1** | 0.675 | **1** | 0.975 | 0.676 | **1** | 0.975 | **1** |
| CharacterTrajectories | 0.984 | **0.997** | 0.994 | **0.997** | 0.985 | 0.988 | 0.99 | 0.964 | 0.969 | 0.99 | 0.964 | 0.969 | 0.989 | 0.989 |
| Cricket | 0.986 | 0.968 | 0.986 | 0.958 | 0.917 | 0.972 | **1** | 0.944 | 0.986 | **1** | 0.944 | 0.986 | **1** | **1** |
| DuckDuckGeese | N/A | N/A | **0.675** | 0.575 | **0.675** | N/A | 0.575 | 0.275 | 0.55 | 0.6 | 0.275 | 0.55 | 0.6 | 0.567 |
| EigenWorms | N/A | N/A | 0.878 | 0.489 | 0.504 | N/A | **0.89** | 0.55 | 0.603 | 0.618 | 0.549 | N/A | 0.619 | N/A |
| Epilepsy | 0.964 | 0.946 | 0.957 | 0.971 | 0.761 | 0.986 | **1** | 0.667 | 0.978 | 0.964 | 0.666 | 0.978 | 0.964 | 0.979 |
| ERing | **0.981** | 0.904 | 0.88 | 0.904 | 0.941 | 0.856 | 0.964 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 | 0.93 |
| EthanolConcentration | 0.399 | 0.373 | 0.236 | 0.323 | 0.373 | 0.323 | **0.43** | 0.293 | 0.304 | 0.323 | 0.293 | N/A | 0.323 | 0.316 |
| FaceDetection | **0.647** | 0.556 | 0.528 | 0.556 | 0.545 | N/A | 0.545 | 0.519 | 0.513 | 0.529 | 0.519 | 0.5 | 0.529 | 0.529 |
| FingerMovements | **0.62** | 0.55 | 0.54 | 0.53 | 0.58 | 0.59 | 0.49 | 0.55 | 0.52 | 0.53 | 0.55 | 0.52 | 0.53 | 0.509 |
| HandMovementD. | **0.554** | 0.365 | 0.27 | 0.378 | 0.365 | 0.324 | 0.365 | 0.279 | 0.306 | 0.231 | 0.278 | 0.306 | 0.231 | 0.224 |
| Handwriting | 0.421 | 0.335 | 0.533 | 0.357 | 0.286 | 0.322 | 0.605 | 0.371 | 0.509 | **0.607** | 0.2 | 0.316 | 0.286 | 0.601 |
| Heartbeat | 0.741 | 0.615 | 0.737 | 0.751 | 0.663 | **0.756** | 0.727 | 0.62 | 0.659 | 0.717 | 0.619 | 0.658 | 0.717 | 0.571 |
| InsectWingbeat | N/A | N/A | 0.16 | **0.208** | 0.167 | N/A | N/A | 0.128 | N/A | 0.115 | 0.128 | N/A | N/A | N/A |
| JapaneseVowels | 0.965 | 0.924 | **0.989** | 0.965 | 0.976 | 0.973 | 0.973 | 0.924 | 0.959 | 0.949 | 0.924 | 0.959 | 0.949 | 0.959 |
| Libras | 0.849 | 0.834 | 0.867 | 0.85 | 0.856 | 0.89 | 0.878 | 0.833 | **0.894** | 0.872 | 0.833 | **0.894** | 0.87 | 0.879 |
| LSST | 0.582 | 0.568 | 0.558 | 0.568 | 0.373 | **0.674** | 0.59 | 0.456 | 0.575 | 0.551 | 0.456 | 0.575 | 0.551 | 0.551 |
| MotorImagery | **0.59** | **0.59** | 0.54 | **0.59** | 0.51 | N/A | 0.51 | 0.39 | N/A | 0.5 | 0.51 | N/A | 0.5 | 0.5 |
| N/ATOPS | 0.922 | 0.87 | 0.944 | 0.939 | 0.889 | **0.956** | 0.87 | 0.86 | 0.85 | 0.883 | 0.85 | 0.85 | 0.883 | 0.883 |
| PEMS-SF | **0.803** | 0.744 | 0.688 | 0.751 | 0.699 | N/A | N/A | 0.705 | 0.734 | 0.711 | 0.705 | 0.734 | 0.711 | 0.73 |
| PenDigits | 0.98 | 0.98 | **0.983** | 0.98 | 0.978 | 0.975 | 0.948 | 0.973 | 0.939 | 0.977 | 0.973 | 0.939 | 0.977 | 0.977 |
| Phoneme | 0.177 | 0.19 | **0.246** | 0.175 | 0.11 | 0.19 | 0.19 | 0.104 | 0.151 | 0.151 | 0.104 | 0.151 | 0.151 | 0.151 |
| RacketSports | 0.849 | 0.816 | 0.862 | 0.868 | 0.803 | 0.875 | **0.934** | 0.868 | 0.842 | 0.803 | 0.868 | 0.842 | 0.803 | 0.858 |
| SelfRegulationSCP1 | **0.887** | 0.874 | 0.771 | 0.739 | 0.874 | 0.734 | 0.71 | 0.771 | 0.765 | 0.775 | 0.771 | 0.765 | 0.775 | 0.786 |
| SelfRegulationSCP2 | **0.567** | 0.539 | **0.556** | 0.55 | 0.472 | N/A | 0.46 | 0.483 | 0.533 | 0.539 | 0.483 | 0.533 | 0.539 | 0.539 |
| SpokenArabicDigits | 0.979 | 0.967 | 0.956 | 0.983 | **0.99** | 0.982 | 0.982 | 0.967 | 0.96 | 0.963 | 0.967 | 0.959 | 0.963 | 0.963 |
| StandWalkJump | **0.533** | 0.4 | 0.4 | 0.4 | 0.067 | 0.2 | 0.333 | 0.2 | 0.333 | 0.2 | 0.2 | 0.333 | 0.2 | 0.333 |
| UWaveGestureLibrary | 0.897 | 0.869 | 0.884 | 0.894 | 0.891 | 0.8 | **0.916** | 0.881 | 0.868 | 0.903 | 0.81 | 0.868 | 0.903 | 0.9 |
| Avg. Rank | 3.85 | 6.19 | 5.9 | 4.73 | 7.33 | 5.45 | 4.66 | 9.3 | 7.43 | 6.37 | 9.37 | 7.88 | 6.83 | 6.21 |
| Wins (Ties) | **11** | **3** | 6 | 5 | 2 | 5 | 8 | 0 | 2 | 2 | 0 | 2 | 1 | 2 |

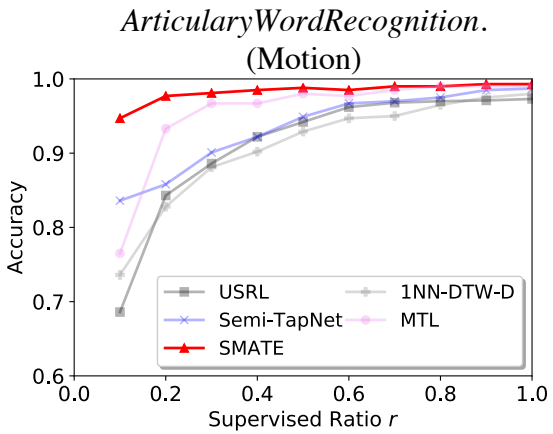
# RQ2. Semi-supervised classification performance

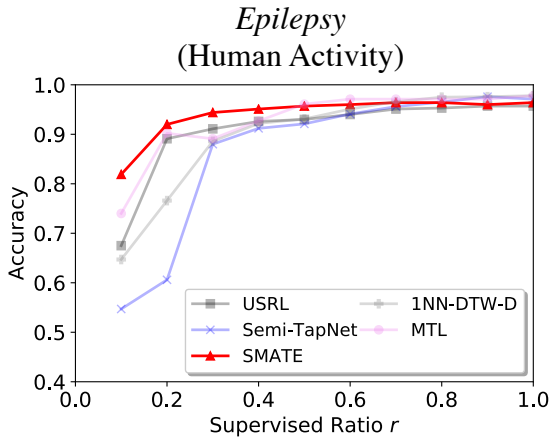
- **Four datasets** on different domains from UEA Archive[1]
  - *ArticularyWordR*. (Motion)
  - *Epilepsy* (Human Activity)
  - *Heartbeat* (Audio Spectra)
  - *SelfRegulationSCP1* (EEG/MEG)
- Baselines:
  - **1NN-DTW-D** [Chen et al., KDD'13]
    - Initially designed for UTS
    - We adjust the distance measure with $DTW_D$ which is designed for MTS
  - **USRL** [Franceschi et al., NeurIPS'19]: SVM on unsupervised representation
  - **Semi-TapNets** [Zhang et al., AAAI'20]: Learning unlabeled samples via Attentional Prototype Network
  - **MTL** [Javed et al., PAKDD'20]: Multi-task learning with self-supervised features from forecasting task

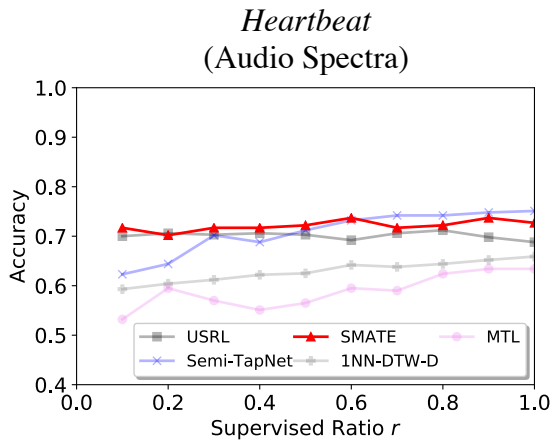

1. www.timeseriesclassification.com

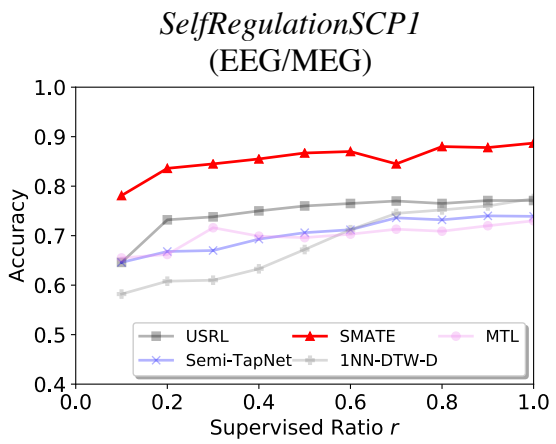
# RQ2. Semi-supervised classification performance

- Semi-supervised representation learning
  - SVM on the learned representation




SMATE performs generally the best among all semi-supervised models, especially under weak supervision

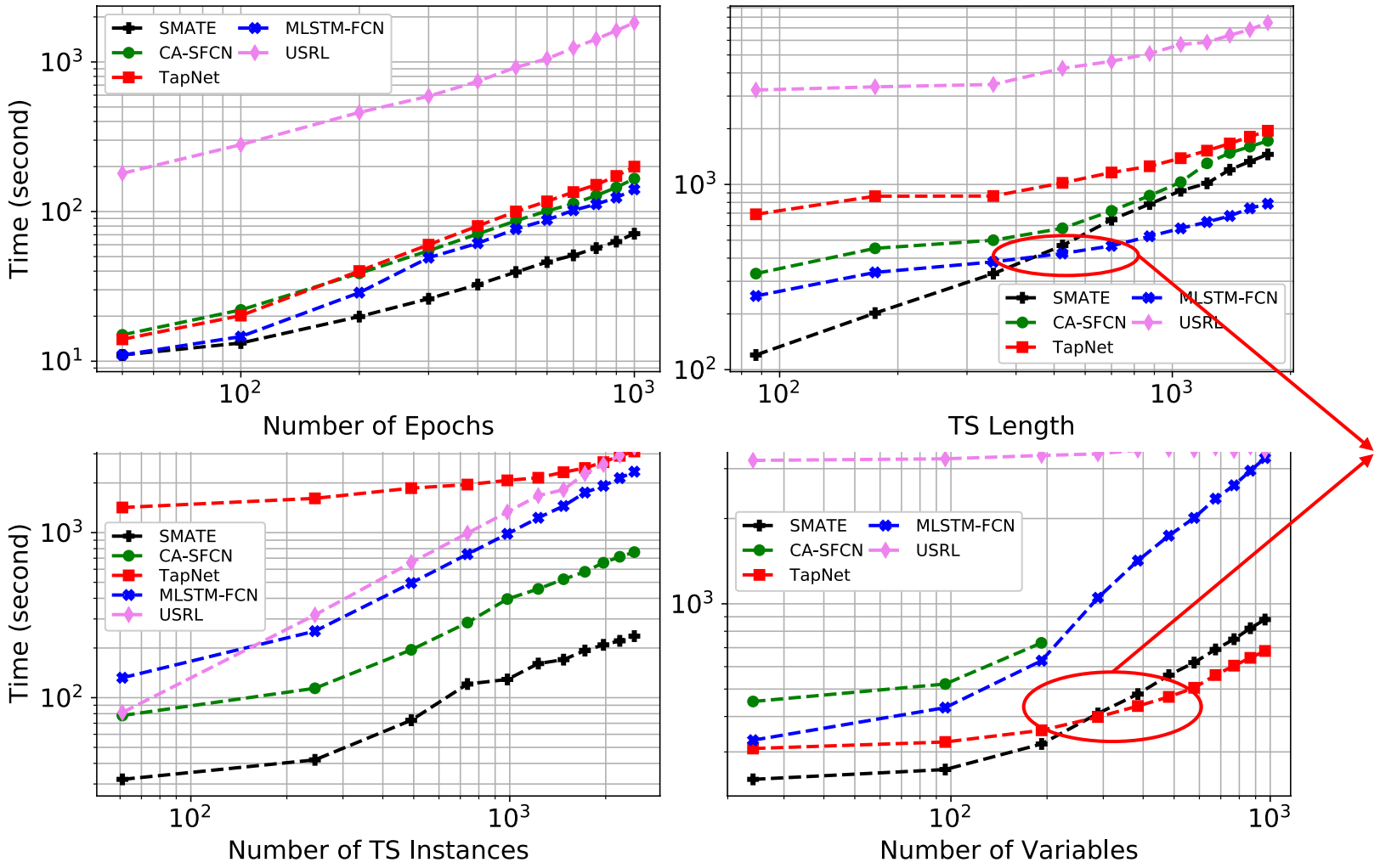
# RQ3. Model efficiency

- Datasets & factors

| Factors | Dataset | (N, M, T) |
|---|---|---|
| Number of training epochs | *ArticularyWordRecognition* | (275, 9, 144) |
| TS length (**T**) | *EthanolConcentration* | (261, 3, **1751**) |
| Number of TS instance (**N**) | *LSST* | (**2459**, 6, 36) |
| Number of variables (**M**) | *PEMS-SF* | (267, **963**, 144) |

- Compare with the Deep Learning models, tested on a single Tesla V100-32Go GPU
  - MLSTM-FCNs [Karim et al., Neural Networks'19]
  - USRL [Franceschi et al., NeurIPS'19]
  - TapNet [Zhang et al., AAAI'20]
  - CA-SFCN [Hao et al. IJCAI'20]

# RQ3. Model efficiency



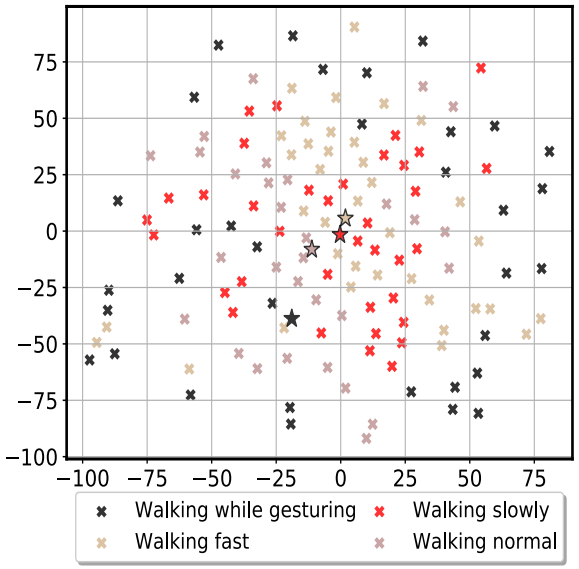

**SMATE** outperforms its 4 competitors in most cases

*Exceptions:*

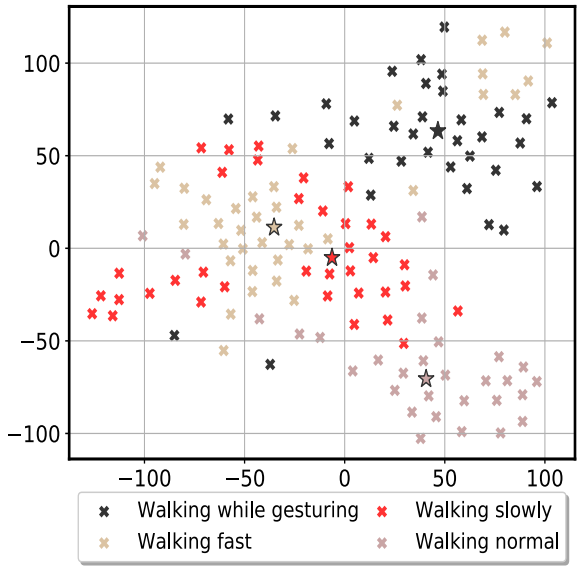
- *on **very long MTS** -> MLSTM-FCN faster*
- *on **MTS with high number of variables** -> TapNet faster*

# RQ4. Interpretation over the representation space

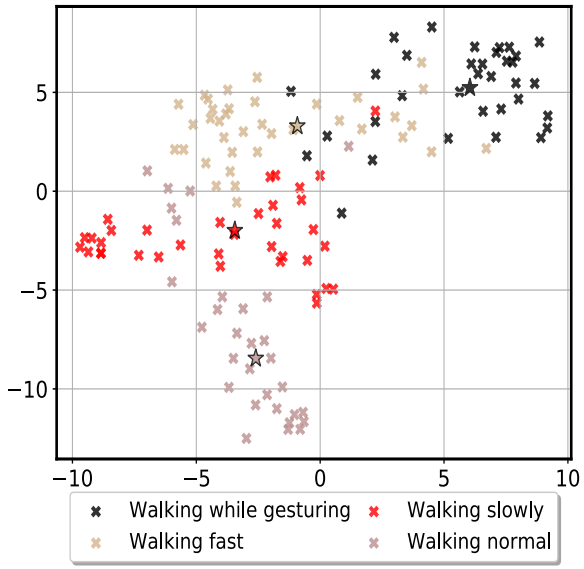
- Interpretable for the effect of the weak supervision
- Interpretable for the classification results



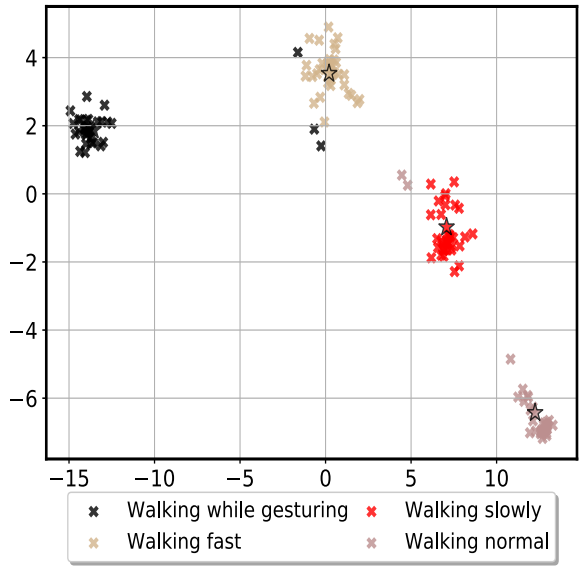

**(a) No regularization**




**(b) Regularization step 1**: supervised initialization




**(c) Regularization step 2**: supervised adjustment

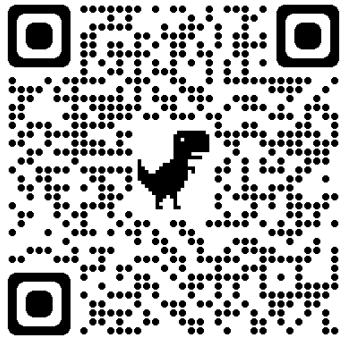



**(d) Regularization step 3**: unsupervised adjustment

*The t-SNE plot of* the data representations (*Epilepsy*);
Supervised ratio = 0.1; Class centroids are marked by ★

# SMATE - Conclusion

- SMATE allows learning a validated Spatio-temporal representation on MTS
  - Spatial Modeling Block (SMB): dynamic spatial interactions
- SMATE allows an efficient representation learning and classification for MTS
- SMATE learns an interpretable representation for:
  - The effect of the weak supervision
  - The classification results
- SMATE allows weak supervision on the embedding space
- Limitations
  - Efficiency problem on the TS that is extra-long & with extra huge variable numbers



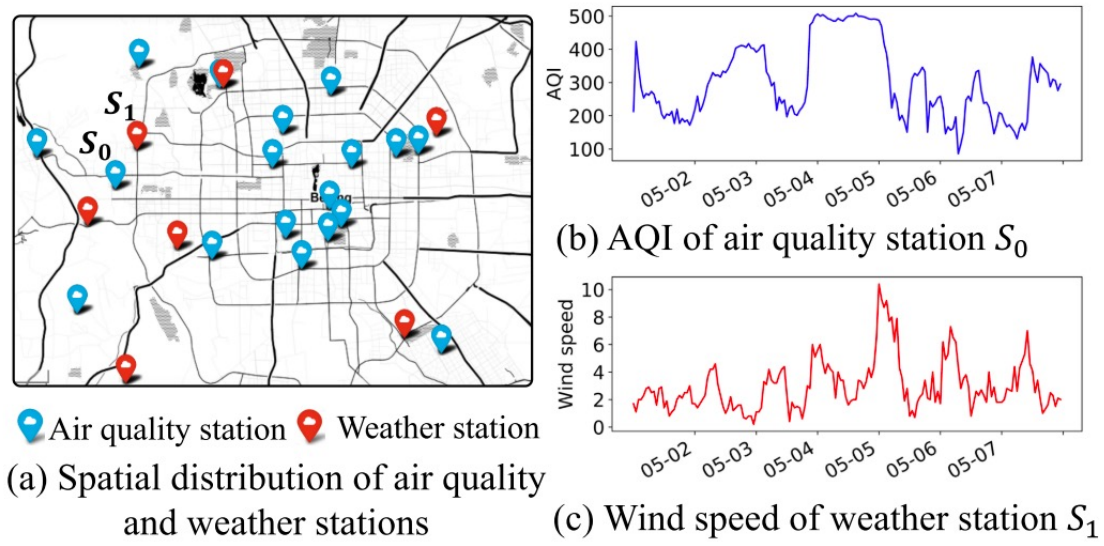
Github page

# Outline

- Introduction
- Background
  - Time series mining
  - Time series representation
- ISMAP: Dynamic Feature Learning on Time Series Stream
- SMATE: Semi-supervised Learning on Multivariate Time Series
- GCN-M: Geo-located Time Series Forecasting with Missing Values
- Conclusion and perspectives

# Context & definitions

- Time series in *Smart City* context:
  - Sensors with fixed spatial locations

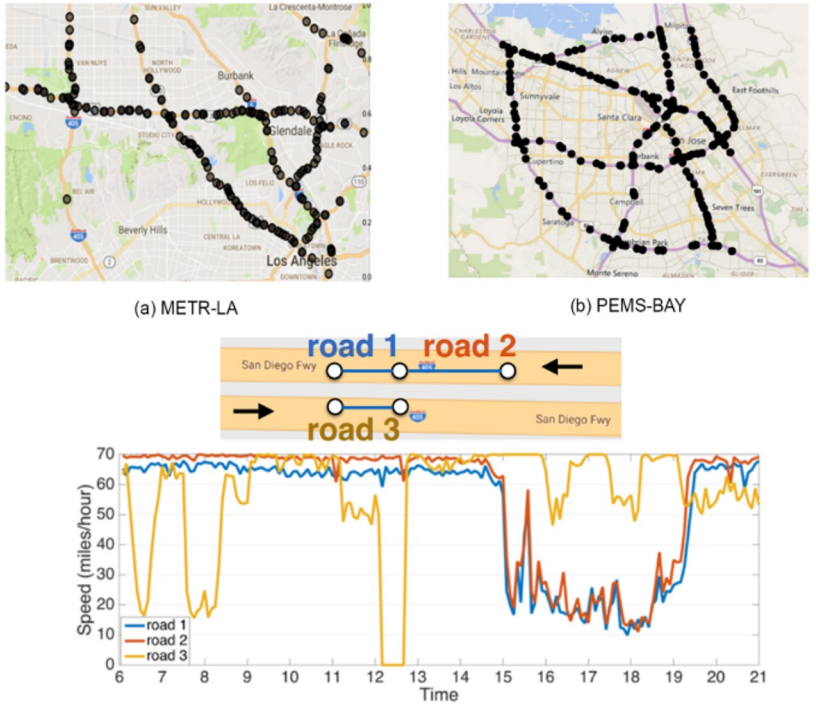



(a) Spatial distribution of air quality and weather stations

(b) AQI of air quality station $S_0$

(c) Wind speed of weather station $S_1$

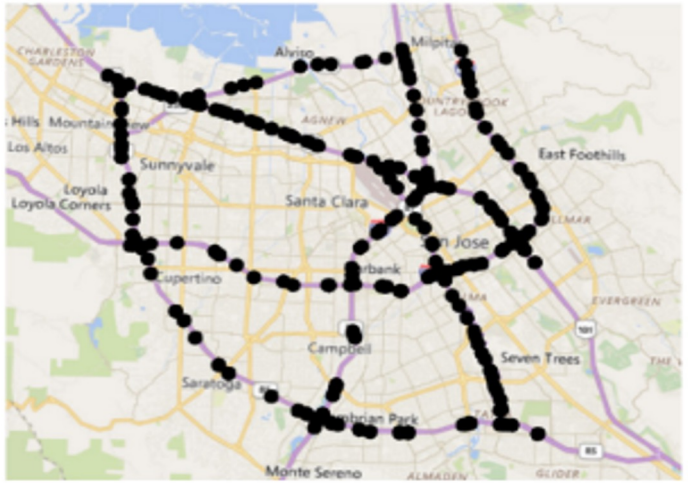
Air quality & weather forecasting [Han et al., AAAI'21]

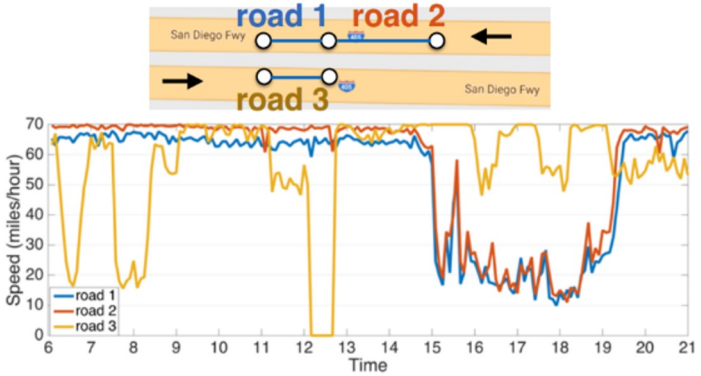



Traffic speed forecasting [Li et al., ICLR'18]

# Context & definitions

- Geo-located time series (GTS)
  - Sensor network $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$
    - $\mathcal{V}$: a set of geo-located nodes
    - $\mathcal{E}$: a set of edges connecting the nodes
  - Sensor data $\mathcal{X} = \{\mathbf{X}_t\}_{t=1}^{T} \in \mathcal{R}^{N \times F \times T}$
    - $N$ : number of spatial nodes
    - $F$ : number of features in each node
    - $T$ : number of timstamps
- Our application context: Traffic forecasting

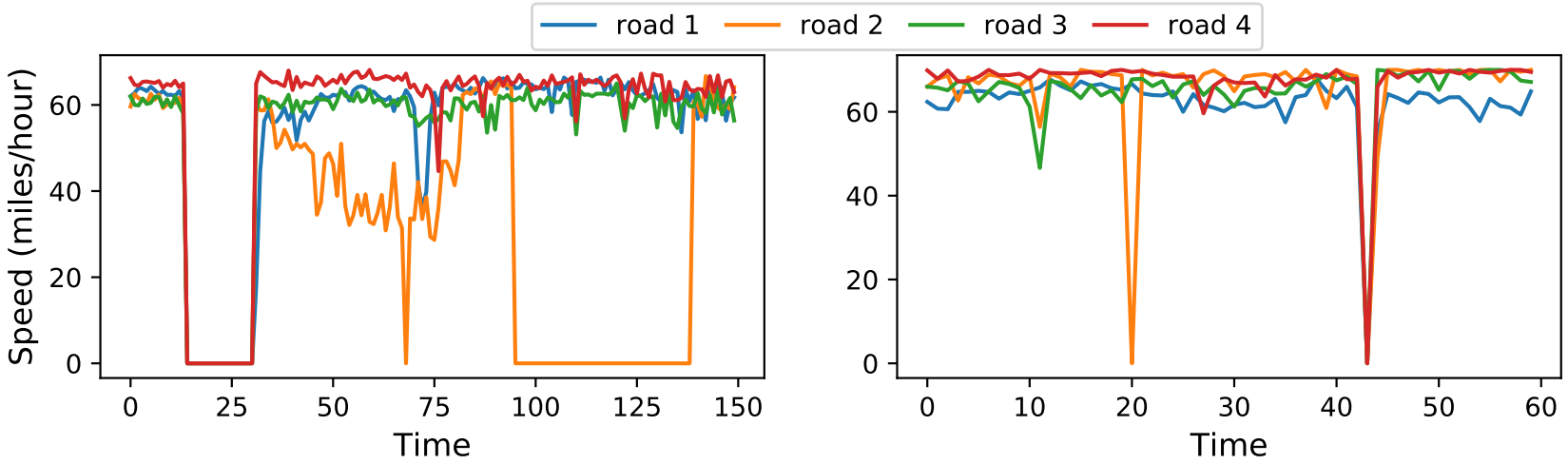



Traffic speed data from PEMS-BAY dataset [Li et al., ICLR'18]

# Problem statement

- Complex scenarios of missing values
  - Temporal: long-range & short-range missing
  - Spatial: partial & entire network missing
  - Hinder the inter-relationship learning in GTS[1]




(a) long-range missing on entire & partial network

(b) short-range missing on partial & entire network

Traffic speed data from METR-LA dataset



1. LOPEZ, Andrés Ladino. Traffic state estimation and prediction in freeways and urban networks. 2018. Thèse de doctorat. Université Grenoble Alpes.

# Objectives

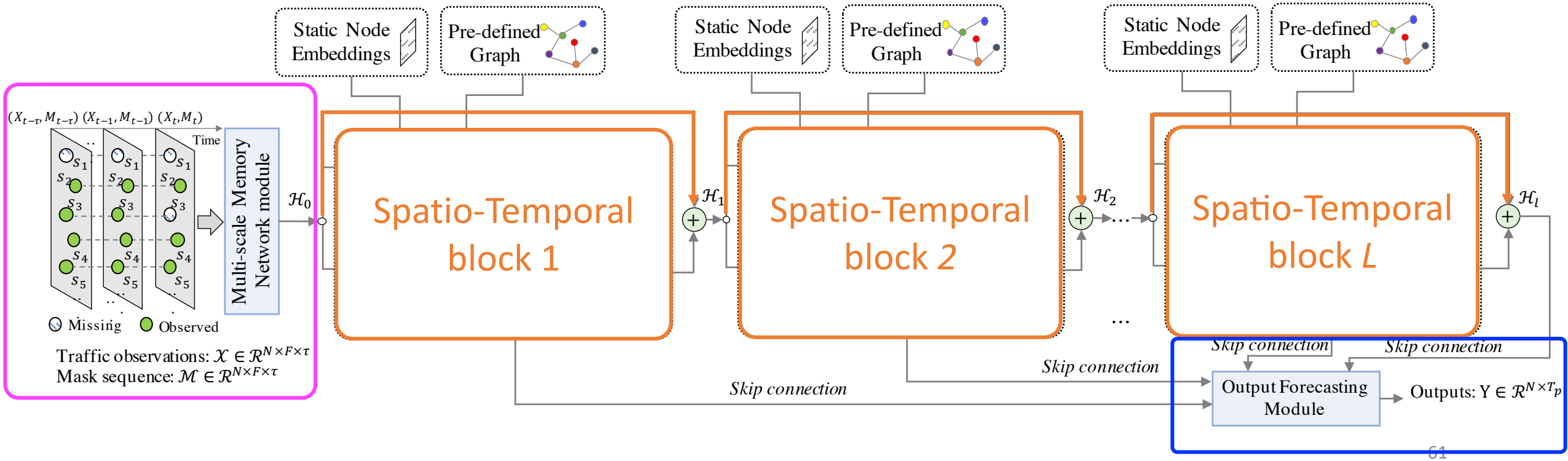
- Complex missing values handling
  - On short & long temporal ranges
  - On partial & entire spatial network
- Spatio-temporal modeling of the inter-relationships in GTS
  - Temporal relationships
  - Spatial relationships
- One-step processing
  - Jointly modeling the Spatio-temporal patterns and complex missing values
- Mainly designed for Traffic Forecasting task
  - Predicting future traffic situations based on the past

# Related work

- Two-step processing
  - Isolate missing-value processing and traffic forecasting
  - e.g., imputation-based models [Yoon et al., NeurIPS'19]
  - x The *general* techniques usually perform worse than the *task-specific* techniques
- One-step processing
  - Jointly model missing values and traffic forecasting
  - e.g., GRU-D [Che et al., Sci. Rep.'18], LSTM-M [Tian et al., Neurocomputing'18], SGMN [Cui et al., Transp. Res. Part C Emerg.'20], LGnet [Tang et al., AAAI'20]
  - x Less considerations on complex missing values and Spatio-temporal patterns
- General traffic forecasting models
  - Ignore missing values during model's optimization
  - e.g., DCRNN [Li et al., ICLR'18], STGCN [Yu et al., IJCAI'18], Graph-Wavenet [Wu et al., IJCAI'19], AGCRN [Bai et al., NeurIPS'20], GTS [Shang et al., ICLR'20], MTGNN [Wu et al., KDD'20]
  - x Ignoring missing values hinders the inter-relationship learning in GTS

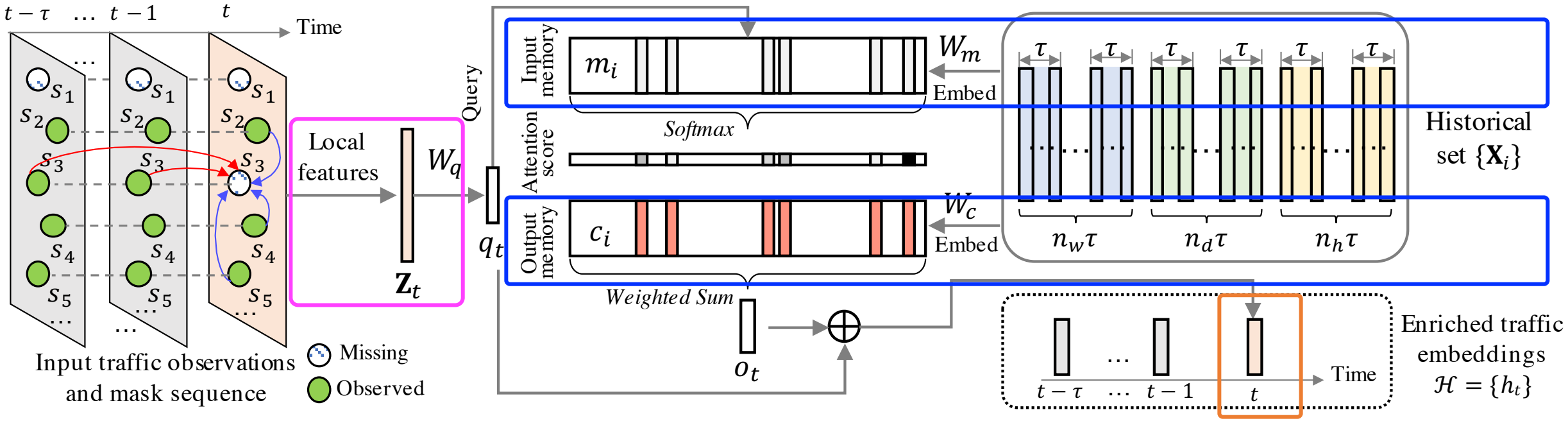
# Proposal - GCN-M

- Graph Convolutional Networks for Traffic Forecasting with Missing Values
  - Multi-scale Memory Network: complex missing value modeling
  - *L* Spatio-Temporal Blocks *(residual connections)*: Spatio-temporal pattern modeling
  - Output Forecasting Module: forecasting results



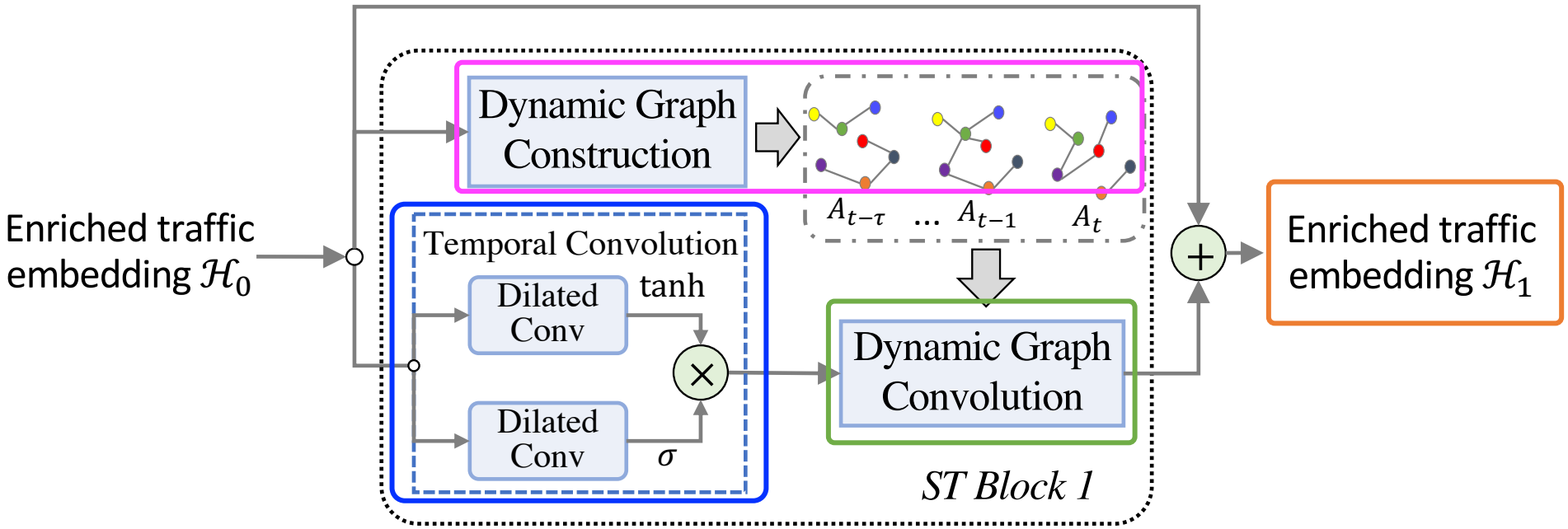
# Multi-scale Memory Network

- Enriched embeddings with local-global features
  - Local statistical features (Keys)
  - Global historical patterns (Memory components)
  - Combine local-global features: use Keys to query the Memory components



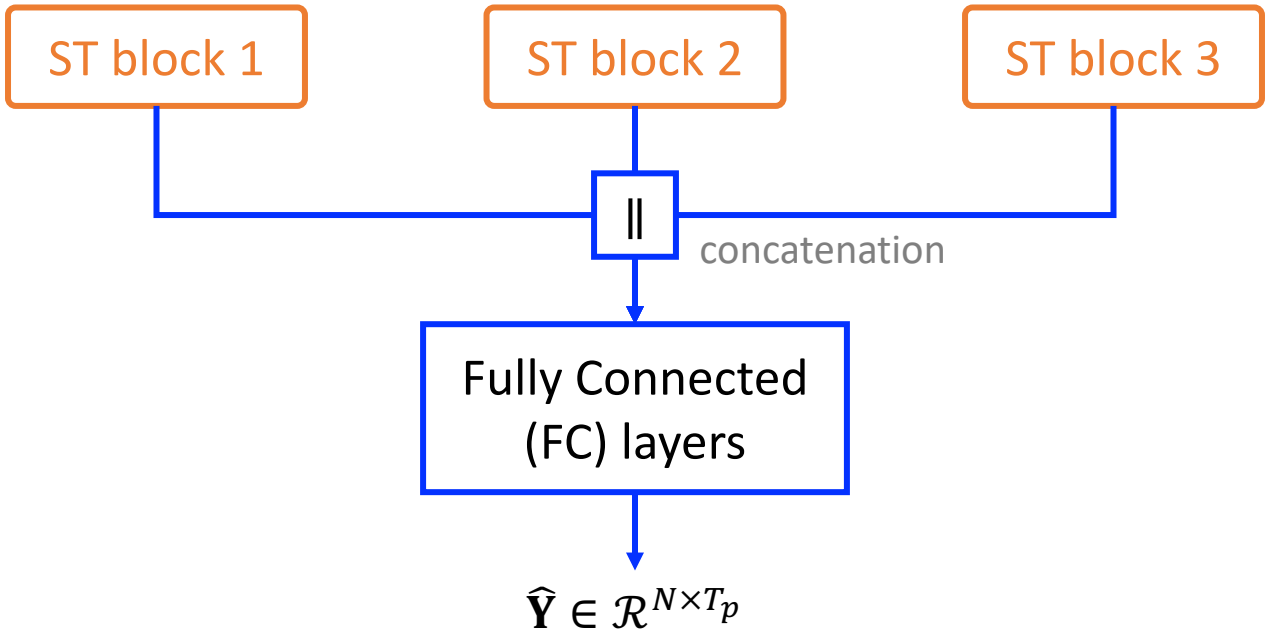
# Spatio-Temporal Block

- Enriched embeddings with Spatio-temporal features
  - Dynamic graph construction : dynamic spatial relationships
  - Temporal convolution: temporal features
  - Dynamic graph convolution: re-weighting temporal features with spatial relationships



# Output Forecasting Module

- Combine features from each ST block
  - $O = (\mathbf{h}_0 W_s^0 + b_s^0) \,\|\, \cdots \,\|\, (\mathbf{h}_i W_s^i + b_s^i) \,\|\, \cdots \,\|\, (\mathbf{h}_{l-1} W_s^{l-1} + b_s^{l-1}) \,\|\, (\mathcal{H}_l W_s^l + b_s^l)$
- Project the concatenated features into the desired output dimension
  - $\widehat{\mathbf{Y}} = W_{fc}^2 (W_{fc}^1 O + b_{fc}^1) + b_{fc}^2 \in \mathcal{R}^{N \times T_p}$
  - Loss function: mean absolute error (MAE), $\mathrm{L} = \frac{1}{N T_p} \sum_{n=1}^{N} \sum_{t=1}^{T_p} \left| \widehat{\mathbf{Y}}_t^n - \mathbf{Y}_t^n \right|$



# Experiments

Research Questions:

- RQ1. Performance on complete datasets
  - How well GCN-M performs on complete traffic datasets?
- RQ2. Complex scenarios of missing values
  - How successful is our model in forecasting traffic data considering the complex missing values?
- RQ3. Dynamic graph modeling
  - How our method performs on dynamic graph modeling considering the missing values?

# Experiments

- Datasets
  - Use recent τ= 12 steps as input to predict the next $T_p \in \{3, 6, 12\}$ steps
  - Artificially mask raw data for simulating complex missing-value scenarios

| Data | #Nodes | #Edges | Length | Sample Rate | Observations | Zero ratio |
|---|---|---|---|---|---|---|
| PEMS-BAY | 325 | 2369 | 52 116 | 5 mins | 16 937 179 | 0.0031% |
| METR-LA | 207 | 1515 | 34 272 | 5 mins | 6 519 002 | 8.11% |

- Evaluation metrics

$$MAE(\mathbf{Y}, \widehat{\mathbf{Y}}) = \frac{1}{NT_p} \sum_{n=1}^{N} \sum_{t=1}^{T_p} \left| \widehat{\mathbf{Y}}_t^n - \mathbf{Y}_t^n \right|$$

$$RMSE(\mathbf{Y}, \widehat{\mathbf{Y}}) = \sqrt{\frac{1}{NT_p} \sum_{n=1}^{N} \sum_{t=1}^{T_p} \left| \widehat{\mathbf{Y}}_t^n - \mathbf{Y}_t^n \right|^2}$$

$$MAPE(\mathbf{Y}, \widehat{\mathbf{Y}}) = \frac{1}{NT_p} \sum_{n=1}^{N} \sum_{t=1}^{T_p} \left| \frac{\widehat{\mathbf{Y}}_t^n - \mathbf{Y}_t^n}{\mathbf{Y}_t^n} \right|$$

- Baselines
  - Six recent traffic forecasting models (Ignore missing values)
    - DCRNN [Li et al., ICLR'18], STGCN [Yu et al., IJCAI'18], Graph-Wavenet [Wu et al., IJCAI'19], AGCRN [Bai et al., NeurIPS'20], GTS [Shang et al., ICLR'20], MTGNN [Wu et al., KDD'20]
  - Five one-step processing models (Joint modeling)
    - (GRU), GRU-I, GRU-D [Che et al., Sci. Rep.'18], LSTM-M [Tian et al., Neurocomputing'18], LSTM-I, SGMN [Cui et al., Transp. Res. Part C Emerg.'20]

# RQ1. Performance on complete datasets

| PEMS-BAY | Horizon=1 (5 mins) | | | Horizon=3 (15 mins) | | | Horizon=6 (30 mins) | | | Horizon=12 (60 mins) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Models | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| DCRNN | 0.96 | 1.63 | 1.81% | 1.38 | 2.95 | 2.90% | 1.74 | 3.97 | 3.90% | 2.07 | 4.74 | 4.90% |
| STGCN | 0.98 | 1.84 | 1.98% | 1.44 | 2.88 | 3.16% | 1.85 | 3.82 | 4.20% | 2.21 | 4.52 | 5.09% |
| GraphWaveNet | 0.91 | **1.56** | 1.72% | **1.31** | 2.75 | **2.73%** | 1.65 | 3.75 | 3.74% | 1.99 | 4.62 | 4.78% |
| MTGNN | **0.87** | 1.57 | **1.70%** | 1.33 | 2.80 | 2.80% | 1.65 | 3.75 | 3.70% | 1.95 | 4.49 | 4.56% |
| AGCRN | 0.95 | 1.81 | 1.94% | 1.37 | 2.92 | 2.94% | 1.69 | 3.87 | 3.82% | 1.99 | 4.61 | 4.62% |
| GTS | 0.91 | 1.64 | 1.77% | 1.32 | 2.80 | 2.75% | 1.63 | 3.74 | **3.63%** | **1.90** | **4.40** | **4.44%** |
| GRU | 1.29 | 2.46 | 2.54% | 1.89 | 3.53 | 3.98% | 2.27 | 4.24 | 5.02% | 2.65 | 4.90 | 5.92% |
| GRU-I | 1.30 | 2.57 | 2.57% | 1.89 | 3.52 | 3.99% | 2.26 | 4.22 | 4.99% | 2.62 | 4.89 | 5.87% |
| GRU-D | 5.40 | 9.25 | 13.83% | 5.34 | 9.25 | 13.76% | 5.42 | 9.26 | 13.85% | 5.41 | 9.27 | 13.85% |
| LSTM-I | 1.71 | 2.69 | 2.80% | 1.97 | 3.45 | 4.08% | 2.57 | 5.52 | 5.62% | 2.74 | 5.00 | 6.21% |
| LSTM-M | 1.35 | 2.31 | 2.71% | 1.87 | 3.39 | 3.95% | 2.33 | 4.33 | 5.17% | 3.45 | 8.32 | 7.29% |
| SGMN | 0.98 | 1.85 | 1.88% | 1.63 | 3.40 | 3.32% | 2.29 | 4.91 | 4.88% | 3.31 | 6.86 | 7.32% |
| **GCN-M (ours)** | 0.91 | 1.57 | 1.75% | 1.33 | **2.72** | 2.76% | **1.62** | **3.64** | 3.64% | 1.95 | **4.40** | 4.61% |

| METR-LA | Horizon=1 (5 mins) | | | Horizon=3 (15 mins) | | | Horizon=6 (30 mins) | | | Horizon=12 (60 mins) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Models | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| DCRNN | 2.45 | 4.21 | 5.99% | 2.77 | 5.38 | 7.30% | 3.15 | 6.45 | 8.80% | 3.60 | 7.60 | 10.50% |
| STGCN | 2.58 | 4.32 | 6.22% | 3.04 | 5.48 | 8.00% | 3.60 | 6.51 | 9.97% | 4.21 | 7.37 | 11.61% |
| GraphWaveNet | 2.41 | 4.29 | 5.93% | **2.68** | **5.14** | 6.87% | 3.06 | 6.14 | 8.23% | 3.52 | 7.25 | **9.77%** |
| MTGNN | **2.24** | 3.92 | **5.39%** | **2.68** | 5.16 | **6.86%** | **3.05** | 6.16 | 8.19% | **3.50** | **7.24** | 9.83% |
| AGCRN | 2.41 | 4.27 | 6.08% | 2.86 | 5.54 | 7.66% | 3.22 | 6.55 | 8.92% | 3.58 | 7.45 | 10.24% |
| GTS | 2.32 | 4.15 | 6.12% | 2.72 | 5.42 | 7.11% | 3.11 | 6.47 | **7.49%** | 3.52 | 7.49 | 10.07% |
| GRU | 2.83 | 4.56 | 6.78% | 3.48 | 5.80 | 9.02% | 3.97 | 6.74 | 10.72% | 4.65 | 7.86 | 13.00% |
| GRU-I | 2.80 | 4.52 | 6.70% | 3.49 | 5.83 | 9.05% | 3.97 | 6.74 | 10.75% | 4.60 | 7.88 | 12.80% |
| GRU-D | 7.46 | 11.82 | 24.55% | 7.43 | 11.85 | 24.62% | 7.45 | 11.84 | 24.62% | 7.47 | 11.86 | 24.68% |
| LSTM-I | 2.86 | 4.57 | 6.77% | 3.57 | 5.88 | 9.05% | 4.10 | 6.85 | 10.94% | 4.78 | 8.13 | 13.34% |
| LSTM-M | 3.15 | 5.58 | 7.03% | 3.46 | 5.74 | 8.75% | 4.08 | 6.86 | 10.89% | 4.63 | 7.83 | 12.92% |
| SGMN | 3.11 | 6.02 | 7.01% | 4.23 | 8.54 | 9.89% | 5.46 | 10.88 | 13.01% | 7.37 | 13.78 | 17.81% |
| **GCN-M (ours)** | 2.34 | **3.89** | 5.88% | 2.74 | 5.21 | 6.94% | 3.12 | 6.18 | 8.25% | 3.54 | **7.12** | 10.01% |

- **Comparable** performance to recent traffic forecasting models
- **Clear advantage** over one-step processing models

# RQ2. Complex scenarios of missing values

| PEMS-BAY | | Missing Rate = 10% | | | Missing Rate = 20% | | | Missing Rate = 40% | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Models | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| Mix-range missing | DCRNN | 1.81 | 4.01 | 4.15% | 1.91 | 4.16 | 4.31% | 2.02 | 4.36 | 4.52% |
| | STGCN | 1.85 | 4.13 | 4.21% | 1.98 | 4.31 | 4.56% | 2.11 | 4.43 | 4.68% |
| | GraphWaveNet | 1.72 | 3.92 | 3.96% | 1.83 | 4.06 | 4.14% | 1.89 | 4.11 | 4.21% |
| | MTGNN | 1.69 | 3.77 | 3.78% | 1.86 | 4.03 | 4.11% | 1.98 | 4.32 | 4.44% |
| | AGCRN | 1.67 | 3.85 | 3.88% | 1.72 | 3.95 | 3.99% | 1.80 | 4.10 | 4.13% |
| | GTS | 1.70 | 3.96 | 3.92% | 1.75 | 3.98 | 3.89% | 1.79 | 4.09 | 4.09% |
| | GRU | 2.71 | 4.88 | 6.03% | 2.82 | 5.08 | 6.28% | 3.05 | 5.43 | 6.82% |
| | GRU-I | 2.31 | 4.30 | 5.11% | 2.34 | 4.39 | 5.18% | 2.40 | 4.50 | 5.37% |
| | GRU-D | 8.90 | 13.71 | 20.03% | 9.46 | 14.50 | 21.04% | 10.21 | 15.19 | 22.44% |
| | LSTM-I | 2.46 | 4.51 | 5.49% | 2.75 | 5.85 | 6.02% | 3.39 | 9.15 | 6.88% |
| | LSTM-M | 3.86 | 7.06 | 8.93% | 5.19 | 9.71 | 13.15% | 5.27 | 9.74 | 13.29% |
| | SGMN | 7.41 | 10.91 | 13.47% | 9.95 | 13.49 | 17.56% | 13.10 | 16.96 | 22.58% |
| | **GCN-M (ours)** | **1.65** | **3.67** | **3.69%** | **1.66** | **3.72** | **3.62%** | **1.69** | **3.79** | **3.83%** |

| METR-LA | | Missing Rate = 10% | | | Missing Rate = 20% | | | Missing Rate = 40% | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Models | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| Mix-range missing | DCRNN | 3.33 | 6.69 | 9.53% | 3.47 | 6.85 | 9.64% | 3.56 | 6.95 | 9.78% |
| | STGCN | 3.56 | 7.12 | 9.81% | 3.64 | 7.28 | 10.33% | 3.73 | 7.51 | 10.62% |
| | GraphWaveNet | 3.28 | 6.51 | 9.02% | 3.43 | 6.78 | 9.52% | 3.51 | 6.94 | 9.62% |
| | MTGNN | **3.04** | **6.18** | **7.84%** | 3.14 | 6.72 | 9.07% | 3.44 | 6.82 | 9.12% |
| | AGCRN | 3.19 | 6.49 | 8.77% | 3.21 | 6.56 | 8.95% | 3.26 | 6.65 | 8.98% |
| | GTS | 3.12 | 6.51 | 8.61% | 3.22 | 6.61 | 8.84% | 3.34 | 6.72 | 8.86% |
| | GRU | 4.30 | 7.14 | 11.47% | 4.35 | 7.31 | 11.68% | 4.65 | 7.72 | 12.58% |
| | GRU-I | 4.05 | 6.83 | 10.94% | 4.01 | 6.86 | 10.83% | 4.11 | 6.97 | 10.98% |
| | GRU-D | 7.53 | 11.89 | 24.74% | 7.71 | 12.32 | 25.43% | 7.89 | 12.34 | 25.63% |
| | LSTM-I | 4.15 | 6.94 | 11.06% | 4.19 | 7.01 | 11.18% | 4.30 | 7.18 | 11.35% |
| | LSTM-M | 4.40 | 7.38 | 11.91% | 5.14 | 8.77 | 14.92% | 6.02 | 9.92 | 17.88% |
| | SGMN | 9.33 | 14.16 | 20.47% | 11.42 | 15.87 | 24.40% | 13.84 | 18.13 | 28.97% |
| | **GCN-M (ours)** | 3.08 | 6.34 | 8.59% | **3.12** | **6.42** | **8.71%** | **3.23** | **6.50** | **8.76%** |

- Mix-range missing
  - Mask short & long range values on Temporal & Spatial axis
  - **Clear advantage** over recent traffic forecasting models
  - **Clear advantage** over one-step processing models

# RQ3. Dynamic Graph Modeling

- Mix-range missing with missing rate = 40%

| GCN-M variants | Pre-defined graph | Learned graph | Static/dynamic graphs | Contruct dynamic graphs with |
|---|---|---|---|---|
| GCN-M-obs | ✓ | ✓ | dynamic | raw observations |
| GCN-M-adp | ✗ | ✓ | static | - |
| GCN-M-pre | ✓ | ✗ | static | - |
| GCN-M-com | ✓ | ✓ | static | - |
| GCN-M | ✓ | ✓ | dynamic | enriched embeddings |

| | Models | Horizon=3 (15 mins) MAE | RMSE | MAPE | Horizon=6 (30 mins) MAE | RMSE | MAPE | Horizon=12 (60 mins) MAE | RMSE | MAPE |
|---|---|---|---|---|---|---|---|---|---|---|
| PEMS-BAY | GCN-M-obs | 1.63 | 3.48 | 3.53% | 1.93 | 4.16 | 4.31% | 2.25 | 4.81 | 5.10% |
| | GCN-M-adp | 1.53 | 3.11 | 3.14% | 1.82 | 3.93 | 4.03% | 2.14 | 4.72 | 4.92% |
| | GCN-M-pre | 1.61 | 3.27 | 3.21% | 1.87 | 4.05 | 4.11% | 2.18 | 4.74 | 5.03% |
| | GCN-M-com | 1.54 | 3.13 | 3.11% | 1.79 | 3.92 | 3.97% | 2.11 | 4.62 | 3.91% |
| | **GCN-M** | **1.45** | **3.03** | **3.09%** | **1.70** | **3.81** | **3.89%** | **2.06** | **4.64** | **4.86%** |
| METR-LA | GCN-M-obs | 2.97 | 5.68 | 7.71% | 3.31 | 6.57 | 8.78% | 3.71 | 7.54 | 10.07% |
| | GCN-M-adp | 2.84 | 5.51 | 7.44% | 3.19 | 6.41 | 8.59% | 3.68 | 7.34 | 9.96% |
| | GCN-M-pre | 2.92 | 5.56 | 7.64% | 3.23 | 6.42 | 8.63% | 3.72 | 7.42 | 10.04% |
| | GCN-M-com | 2.84 | 5.52 | 7.45% | 3.17 | 6.4 | 8.63% | 3.68 | 7.41 | 9.97% |
| | **GCN-M** | **2.82** | **5.47** | **7.42%** | **3.16** | **6.38** | **8.55%** | **3.58** | **7.31** | **9.92%** |

# Model efficiency

- Moderate efficiency performance
  - Performs better than DCRNN, but worse than others
- Caused by
  - Costly computations on the multi-scale memory networks (attention mechanism)
  - Costly convolutions on dynamic graphs

Training time (second) per epoch (on a single Tesla V100-32Go GPU)

| Models | PEMS-BAY | METR-LA | Models | PEMS-BAY | METR-LA |
|---|---|---|---|---|---|
| DCRNN | 468.22 | 178.23 | GRU | 3.65 | 2.45 |
| STGCN | 55.32 | 27.70 | GRU-I | 4.22 | 3.67 |
| GraphWaveNet | 118.77 | 48.16 | GRU-D | 7.82 | 5.43 |
| MTGNN | 86.20 | 38.70 | LSTM-I | 4.32 | 4.64 |
| AGCRN | 67.40 | 32.9 | LSTM-M | 8.12 | 5.76 |
| GTS | 191.4 | 62.3 | SGMN | 3.45 | 2.38 |
| GCN-M (ours) | 241.69 | 118.65 | - | - | - |

# GCN-M - Conclusion

- GCN-M considers the complex scenarios of missing values (long-range & short-range, partial & entire network missing) in traffic data
- GCN-M models the complex inter-relationships in traffic data
- GCN-M jointly models the Spatio-temporal patterns and missing values in one-step processing
- GCN-M is applicable not only to Traffic Forecasting but also to:
  - Crowd flow forecasting
  - Weather and air pollution forecasting
  - etc.

# Conclusion

- ISMAP: Dynamic Feature Learning on Time Series Stream
  - Time series in streaming context
  - Incremental, adaptive and interpretable Shapelet for online classification task
- SMATE: Semi-supervised Learning on Multivariate Time Series
  - Multivariate time series in label-constraint context
  - Efficient, interpretable deep representations for semi-supervised classification task
- GCN-M: Geo-located Time Series Forecasting with Missing Values
  - Traffic time series in Smart City context with data quality issues
  - Powerful deep representations for traffic forecasting task

# Perspectives

- ISMAP: Dynamic Feature Learning on Time Series Stream
  - Extend univariate time series (UTS) stream to multivariate time series (MTS) stream
  - Multi-dimensional Shapelet extraction on Matrix Profile
- SMATE: Semi-supervised Learning on Multivariate Time Series
  - Apply our proposal in GCN-M (e.g., dynamic GCN) to improve the inter-relationship learning in MTS for classification tasks
  - Optimize the semi-supervised framework via e.g., domain adaptation
- GCN-M: Geo-located Time Series Forecasting with Missing Values
  - Improve the efficiency of GCN-M via recent efficient attention mechanisms or graph tensor decomposition
  - Validate GCN-M in wider contexts, e.g., air pollution forecasting

# Publications

## Journals

1. **J. Zuo**, K. Zeitouni, Y. Taher, S. G. Rodriguez. “GCN-M: Graph Convolutional Networks for Traffic Forecasting with Missing Values”. *Data Mining and Knowledge Discovery (DMKD), Springer (2022)*
2. H. El Hafyani, M. Abboud, **J. Zuo**, K. Zeitouni and Y. Taher. “Learning the Micro-environment from Rich Trajectories in the context of Mobile Crowd Sensing -Application to Air Quality Monitoring”. *Geoinformatica, Springer (2022)*

## International Conferences

1. **J. Zuo**, K. Zeitouni, Y. Taher. “SMATE: Semi-supervised Spatio-Temporal Representation Learning on Multivariate Time Series”, *IEEE International Conference on Data Mining (ICDM'21)*
2. **J. Zuo**, K. Zeitouni, Y. Taher. “Incremental and Adaptive Feature Exploration over Time Series Stream”, *IEEE International Conference on Big Data (IEEE BigData’19)*

## National Conference

1. **J. Zuo**, K. Zeitouni, Y. Taher. “Time Series meet Data Streams: Perspectives of the Interdisciplinary Collision and Applications”. *BDA 2019, Lyon, France*

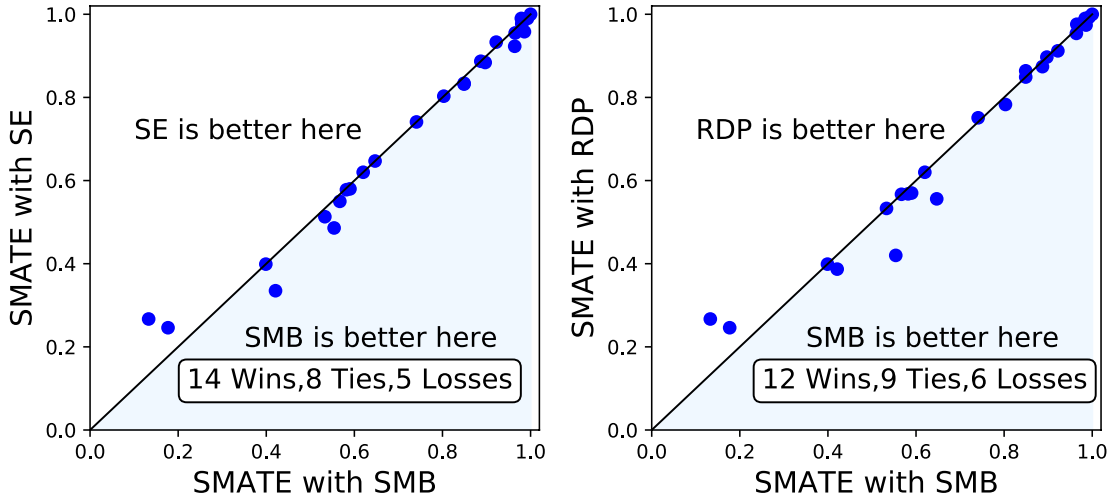
## Workshops & Demos

1. H. El Hafyani, M. Abboud, **J. Zuo**, K. Zeitouni and Y. Taher. “Tell Me What Air You Sense/Breath, I Tell You Where You Are”, *International Symposium on Spatial and Temporal Databases 2021 (SSTD’21), demo.*
2. M. Abboud, H. El Hafyani, **J. Zuo**, K. Zeitouni and Y. Taher. “Micro-environment Recognition in the context of Environmental Crowdsensing”, in *Big Mobility Data Analytics with EDBT 2021 (BMDA’21)*
3. **J. Zuo**, K. Zeitouni, and Y. Taher. “ISETS: Incremental Shapelet Extraction from Time Series Stream”, *ECML-PKDD’19*, demo.
4. **J. Zuo**, K. Zeitouni, and Y. Taher. “Exploring interpretable features for large time series with SE4TeC.”, *EDBT 2019*, demo.

Thank you for your attention.

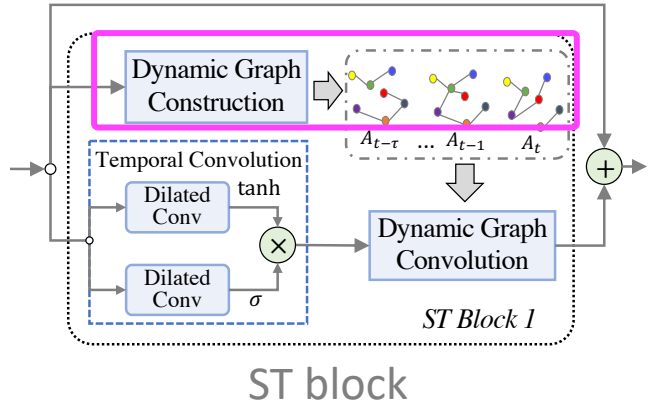
# RQ5: Performance of the Spatial Modeling Block

- **27 datasets from UEA archive** in which SMATE has successfully executed
- SMATE **with SMB** *Versus* SMATE **without SMB**:
  - [**17 Wins** | 8 Ties | 2 Losses]
- **SMB** *Versus* **others**:
  - Squeez-and-Excitation (**SE**) in MLSTM-FCNs [Karim et al., Neural Networks'19]
  - Random Dimension Permutation (**RDP**) in TapNet [Zhang et al., AAAI'20]

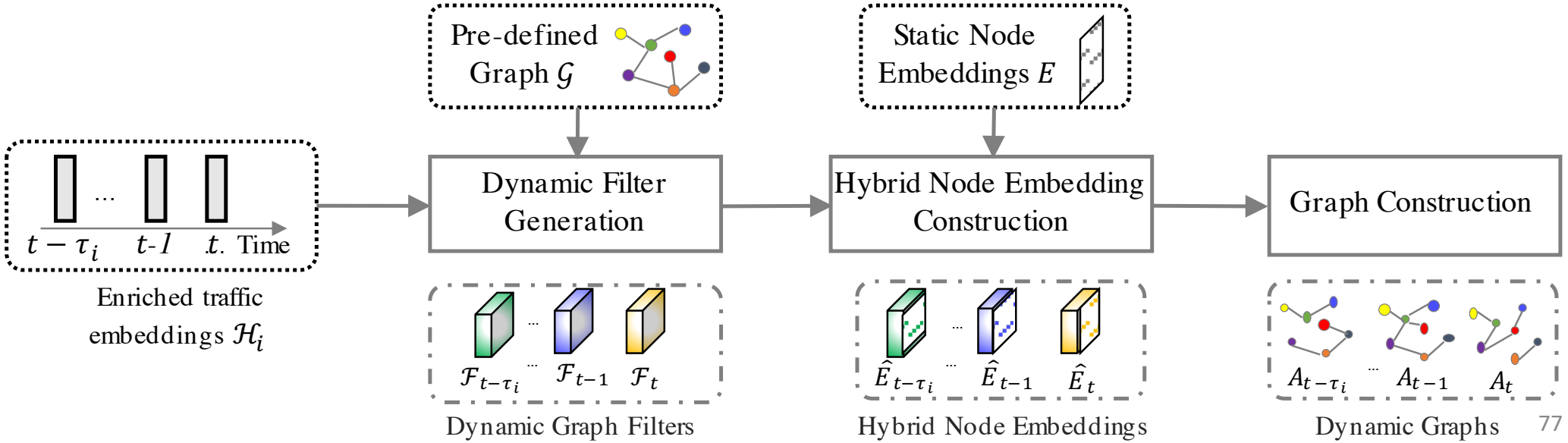


# Dynamic graph construction

- Input
  - Enriched traffic embeddings: observed dynamic node features
  - Pre-defined graph: introduced spatial information
  - Static node embeddings: unobserved static node features
- Output
  - Dynamic graphs at each timestamp

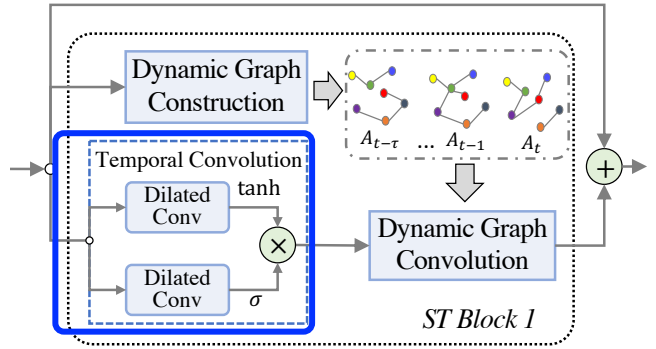



ST block

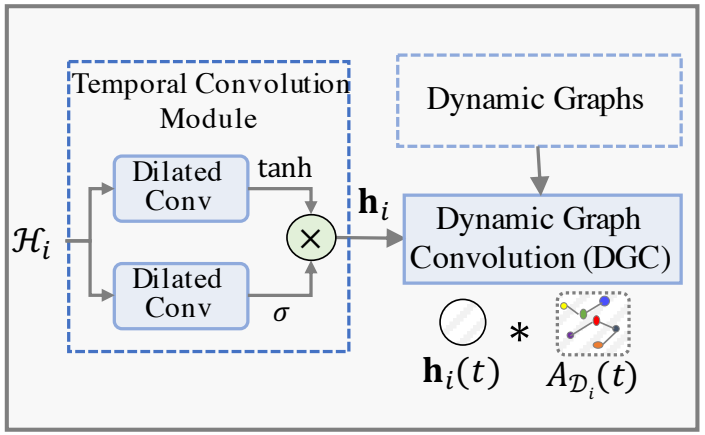


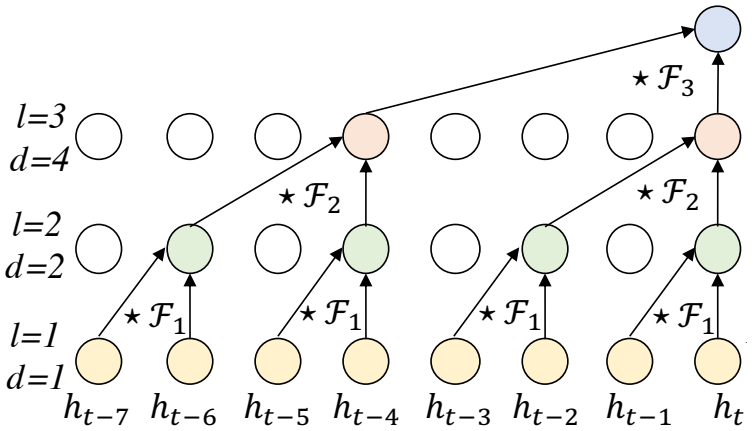
# Temporal convolution

- Extract structural temporal features
  - $\mathbf{h}_i = \tanh(W_{\mathcal{F}} \star \mathcal{H}_i) \odot \sigma(W_{\mathcal{F}'} \star \mathcal{H}_i)$
  - $W_{\mathcal{F}}, W_{\mathcal{F}'}$ : learnable parameters of convolution filters
  - Gating mechanism $\sigma(\cdot)$
    - a Sigmoid activation function which selects structural temporal features




ST block

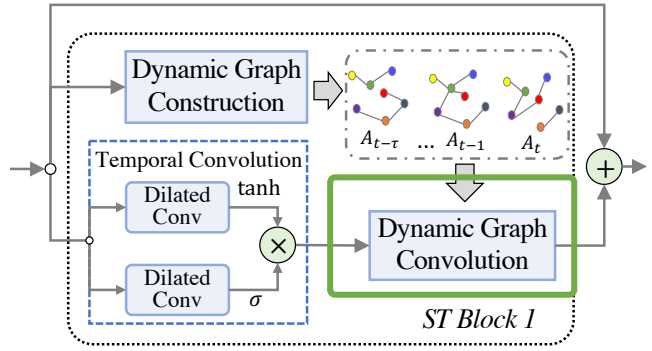


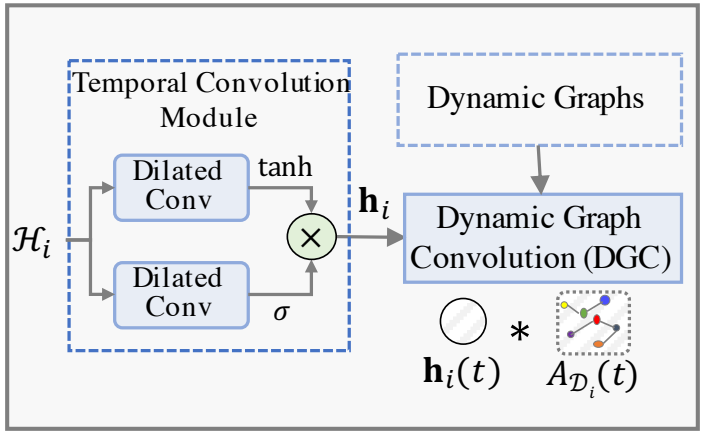

Temporal Dilated Convolution, $k = 2$
(Dilated Conv.)

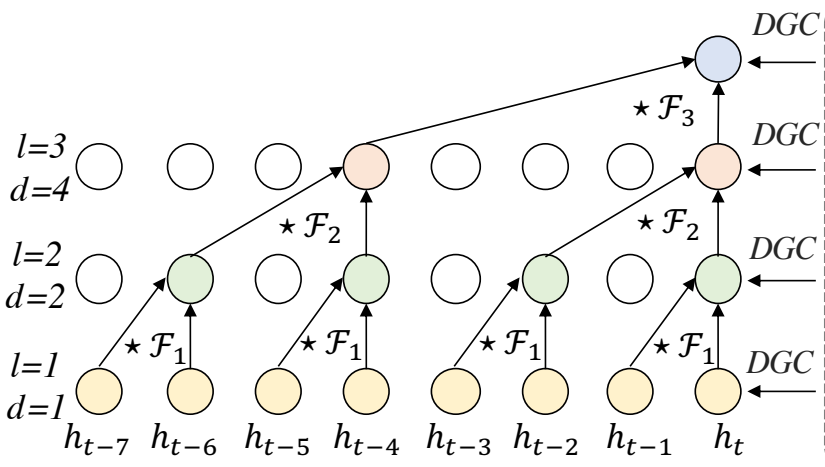
# Dynamic graph convolution

- Aggregate spatial information with temporal features
  - $\mathcal{H}_i'(t) = \sum_{k=0}^{K} (A_i(t))^k \mathbf{h}_i(t) W_k \in \mathcal{R}^{N \times d}$
  - $K$ : diffusion step
  - $A_i(t)$: adjacency matrix (i.e., graph) at time *t*, in the *i-th* ST block
  - $W_k$: learnable parameter matrix

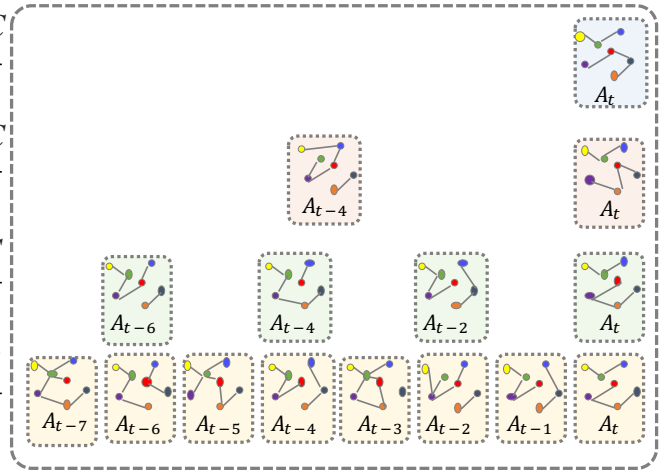



ST block




Temporal Dilated Convolution, $k = 2$
(Dilated Conv.)




Dynamic graphs

# RQ2. Complex scenarios of missing values

| PEMS-BAY | | Missing Rate = 10% | | | Missing Rate = 20% | | | Missing Rate = 40% | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Models | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| Short-range missing | DCRNN | 1.76 | 3.94 | 3.94% | 1.82 | 3.96 | 4.01% | 1.85 | 4.26 | 4.04% |
| | STGCN | 1.82 | 4.11 | 4.25% | 1.91 | 4.18 | 4.41% | 1.97 | 4.33 | 4.42% |
| | GraphWaveNet | 1.69 | 3.79 | 3.81% | 1.74 | 3.75 | 3.75% | 1.79 | 3.87 | 3.90% |
| | MTGNN | **1.58** | **3.42** | **3.33%** | 1.72 | 3.78 | 3.83% | 1.83 | 4.03 | 3.94% |
| | AGCRN | 1.65 | 3.81 | 3.78% | 1.66 | 3.81 | 3.79% | 1.72 | 3.96 | 3.95% |
| | GTS | 1.65 | 3.86 | 3.74% | 1.65 | 3.86 | 3.76% | **1.69** | 3.92 | **3.86%** |
| | GRU | 2.60 | 4.64 | 5.75% | 2.67 | 4.78 | 5.90% | 2.86 | 5.10 | 6.37% |
| | GRU-I | 2.29 | 4.28 | 5.06% | 2.31 | 4.31 | 5.09% | 2.41 | 4.47 | 5.38% |
| | GRU-D | 5.38 | 9.29 | 13.84% | 5.46 | 9.36 | 13.96% | 7.20 | 11.58 | 16.91% |
| | LSTM-I | 2.35 | 4.33 | 5.22% | 2.82 | 6.63 | 6.05% | 3.06 | 7.47 | 6.56% |
| | LSTM-M | 2.47 | 4.55 | 5.50% | 2.56 | 4.70 | 5.74% | 3.34 | 7.09 | 7.68% |
| | SGMN | 2.32 | 4.96 | 4.94% | 2.34 | 5.01 | 5.00% | 2.45 | 5.20 | 5.23% |
| | **GCN-M (ours)** | 1.62 | 3.67 | 3.60% | **1.63** | **3.73** | **3.68%** | 1.75 | **3.81** | 3.90% |

| METR-LA | | Missing Rate = 10% | | | Missing Rate = 20% | | | Missing Rate = 40% | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Models | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| Short-range missing | DCRNN | 3.31 | 6.61 | 9.47% | 3.44 | 6.80 | 9.57% | 3.50 | 6.90 | 9.68% |
| | STGCN | 3.53 | 7.08 | 9.73% | 3.59 | 7.25 | 10.25% | 3.66 | 7.45 | 10.41% |
| | GraphWaveNet | 3.28 | 6.60 | 9.11% | 3.36 | 6.74 | 9.50% | 3.45 | 6.81 | 9.57% |
| | MTGNN | **2.98** | **6.03** | **8.35%** | 3.19 | **6.44** | 8.69% | 3.26 | 6.59 | 9.07% |
| | AGCRN | 3.19 | 6.47 | 8.81% | 3.24 | 6.60 | 9.01% | 3.25 | 6.61 | 9.19% |
| | GTS | 3.08 | 6.40 | 8.59% | **3.14** | 6.52 | **7.58%** | **3.12** | 6.56 | **8.61%** |
| | GRU | 4.20 | 7.09 | 11.27% | 4.27 | 7.16 | 11.42% | 4.45 | 7.41 | 11.94% |
| | GRU-I | 4.02 | 6.83 | 10.89% | 4.03 | 6.88 | 10.83% | 4.09 | 6.91 | 10.88% |
| | GRU-D | 7.50 | 11.87 | 24.69% | 7.45 | 11.86 | 24.66% | 7.53 | 11.91 | 24.76% |
| | LSTM-I | 4.12 | 6.89 | 11.04% | 4.18 | 6.98 | 11.10% | 4.21 | 7.08 | 11.26% |
| | LSTM-M | 4.10 | 6.92 | 10.91% | 4.15 | 6.98 | 11.03% | 4.26 | 7.18 | 11.44% |
| | SGMN | 5.54 | 10.93 | 13.17% | 5.61 | 10.99 | 13.35% | 5.81 | 11.18 | 13.83% |
| | **GCN-M (ours)** | 3.17 | 6.33 | 8.72% | 3.23 | 6.47 | 8.99% | 3.26 | **6.35** | 8.98% |

- Short-range missing
  - **Comparable** performance to recent traffic forecasting models
  - **Clear advantage** over one-step processing models

# RQ2. Complex scenarios of missing values

| PEMS-BAY | | Missing Rate = 10% | | | Missing Rate = 20% | | | Missing Rate = 40% | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Models | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| Long-range missing | DCRNN | 1.83 | 4.07 | 4.22% | 1.96 | 4.22 | 4.42% | 2.07 | 4.45 | 4.67% |
| | STGCN | 1.92 | 4.22 | 4.42% | 2.03 | 4.37 | 4.72% | 2.14 | 4.52 | 4.76% |
| | GraphWaveNet | 1.74 | 3.96 | 4.03% | 1.87 | 4.09 | 4.18% | 1.94 | 4.21 | 4.33% |
| | MTGNN | **1.65** | **3.68** | **3.72%** | 1.89 | 4.01 | 4.17% | 2.01 | 4.42 | 4.61% |
| | AGCRN | 1.72 | 3.78 | 3.94% | 1.84 | 4.11 | 4.13% | 1.90 | 4.18 | 4.31% |
| | GTS | 1.68 | 3.86 | 3.91% | 1.78 | 4.12 | 4.97% | 1.88 | 4.17 | 4.22% |
| | GRU | 2.93 | 5.12 | 6.32% | 3.06 | 5.31 | 6.63% | 3.35 | 5.78 | 7.03% |
| | GRU-I | 2.52 | 4.51 | 5.33% | 2.53 | 4.57 | 5.73% | 2.71 | 4.82 | 5.51% |
| | GRU-D | 9.33 | 14.51 | 22.31% | 9.89 | 13.94 | 22.86% | 11.07 | 15.88 | 23.13% |
| | LSTM-I | 2.65 | 4.65 | 5.88% | 3.13 | 6.35 | 6.82% | 3.62 | 9.53 | 7.12% |
| | LSTM-M | 3.93 | 7.25 | 9.17% | 5.45 | 10.06 | 13.67% | 5.57 | 10.12 | 14.59% |
| | SGMN | 8.86 | 12.57 | 14.54% | 11.45 | 14.56 | 18.31% | 14.62 | 17.23 | 23.13% |
| | **GCN-M (ours)** | 1.70 | 3.75 | 3.74% | **1.73** | **3.88** | **3.92%** | **1.79** | **4.07** | **4.14%** |

| METR-LA | | Missing Rate = 10% | | | Missing Rate = 20% | | | Missing Rate = 40% | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Models | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| Long-range missing | DCRNN | 3.46 | 6.78 | 9.62% | 3.54 | 6.96 | 9.75% | 3.62 | 7.02 | 9.89% |
| | STGCN | 3.71 | 7.20 | 9.91% | 3.76 | 7.39 | 10.42% | 3.88 | 7.66 | 10.67% |
| | GraphWaveNet | 3.43 | 6.64 | 9.07% | 3.57 | 6.92 | 9.62% | 3.61 | 7.03 | 10.71% |
| | MTGNN | 3.19 | **6.32** | **8.48%** | 3.39 | 6.85 | 9.21% | 3.50 | 6.95 | 9.74% |
| | AGCRN | 3.31 | 6.54 | 8.94% | 3.33 | 6.68 | 8.98% | 3.33 | 6.78 | 9.45% |
| | GTS | 3.25 | 6.61 | 8.93% | 3.29 | 6.74 | 8.85% | 3.46 | 6.86 | 9.37% |
| | GRU | 4.37 | 7.28 | 12.54% | 4.47 | 7.44 | 11.72% | 4.76 | 7.81 | 12.71% |
| | GRU-I | 4.20 | 6.91 | 11.78% | 4.09 | 6.97 | 15.42% | 4.21 | 7.03 | 11.43% |
| | GRU-D | 7.59 | 11.94 | 24.72% | 7.82 | 12.46 | 25.67% | 7.96 | 12.45 | 26.12% |
| | LSTM-I | 4.20 | 7.09 | 11.08% | 4.25 | 7.12 | 11.32% | 4.36 | 7.32 | 11.56% |
| | LSTM-M | 4.53 | 7.47 | 11.88% | 5.21 | 8.84 | 15.34% | 6.08 | 10.02 | 18.13% |
| | SGMN | 9.47 | 14.30 | 20.72% | 11.49 | 16.01 | 24.55% | 13.97 | 18.24 | 29.10% |
| | **GCN-M (ours)** | **3.18** | 6.39 | 8.71% | **3.23** | **6.56** | **8.78%** | **3.27** | **6.68** | **9.12%** |

- Long-range missing
  - **Clear advantage** over recent traffic forecasting models for a high missing rate
  - **Clear advantage** over one-step processing models